Australian Government
Department of Defence
Defence Science and
Technology Organisation

An Overview of Conceptual Frameworks

*Dale Lambert, Adam Saulwick, Chris Nowak, Martin Oxenhan
and Damien O'Dea*

Command, Control, Communications and Intelligence Division
Defence Science and Technology Organisation

DSTO-TR-2163

ABSTRACT

This document provides a short overview of the semantic frameworks that are applicable to information fusion applications. The frameworks considered include: STE, NSM, CLCE, DOLCE, OpenCYC, Mephisto, SUMO, BFO, and JC3IEDM.

Published by

Command, Control, Communications and Intelligence Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: (08) 8259 5555
Fax: (08) 8259 6567

AR-014-249
May 2008

An Overview of Conceptual Frameworks

Executive Summary

Information fusion refers to the process whereby machines utilise one or more data sources over time to assemble a representation of aspects of interest in an environment. Historically, the data sources were confined to conventional sensors. However, the rise of terrorism and network centric warfare over the last decade has expanded the scope of information fusion beyond conventional sensor data, with the aspects of interest in the environment now also including biographical, economic, social, transport and telecommunications, geographic, military, political and technical information. One challenge introduced by this transition is how to represent these differing types of information within a machine so that the machine is able to produce meaningful information to end users.

This document briefly surveys some extant frameworks for meaningfully representing information within a machine. It has been prepared by the Defence Science and Technology Organisation (DSTO) in the context of a programme agreement between the DSTO and National ICT Australia (NICTA). That programme combines considerations of natural language frameworks, conceptual frameworks, formal frameworks and computational frameworks, under the view that such a combination is required to facilitate a future user's information fusion needs. This document briefly surveys the STE, NSM, CLCE, DOLCE, OpenCYC, Mephisto, SUMO, BFO, and JC3IEDM frameworks and concludes that a combination of these frameworks will be required to address the combined natural language, conceptual, formal and computational ambitions of the joint programme.

Authors

Dr Dale Lambert
Command, Control, Communications & Intelligence Division

Dr. Dale Lambert is Head of the DSTO Fusion for Situation Awareness Initiative. The DSTO Fusion for Situation Awareness Initiative is a cross divisional long-range research programme directed at understanding the defence and national security information fusion requirement in the network centric environment, with a view to consolidating existing, and innovating new, science and technology to address these requirements.

Dr Adam Saulwick
Command, Control, Communications & Intelligence Division

Dr Adam Saulwick is a research scientist in Human Interaction Capabilities Discipline. His current research focuses on theoretical and practical issues for modelling natural language with particular regard to lexical and constructional semantics, and the interface of morphology, syntax, semantics and discourse. He also conducts research in knowledge representation for information fusion. His post-doctoral research focused on the formal representation of linguistic concepts in ontologies for data integration. His doctoral dissertation was a fieldwork-and-corpus-based grammatical description, text collection and dictionary of a polysynthetic, highly endangered Australian Aboriginal language.

Dr Chris Nowak
Command, Control, Communications & Intelligence Division

Dr. Chris Nowak is Deputy Head of the DSTO Fusion for Situation Awareness Initiative. He received his Ph.D. degree from the University of Adelaide in 1998. He began at DSTO in 1998 as a research scientist in the Information Technology Division working on defence-related knowledge representation activities. In April, 2007, he joined the Fusion for Situation Awareness Initiative, which is attached to C3I Division. His research interests include knowledge representation and reasoning, formal ontologies and logics, and information fusion.

Dr Martin Oxenham

Command, Control, Communications & Intelligence Division

Dr. Martin Oxenham is currently Deputy Head of the Fusion for Situation Awareness Initiative within the Defence Science and Technology Organisation (DSTO). He joined DSTO from the Department of Pure Mathematics at the University of Adelaide where he received a Ph.D. in 1992 in the field of finite geometry.

Martin began at DSTO in 1994 as a research scientist in the High Frequency Radar Division at DSTO working on defence-related data fusion activities. Since then, he has conducted research in the areas of target identification, situation assessment, threat evaluation and weapons allocation, mostly in support of the acquisition of the Wedgetail Early Warning and Control capability under project AIR 5077. In 2002, he was promoted to senior research scientist within Intelligence, Surveillance and Reconnaissance Division. In April, 2007, he joined the Fusion for Situation Awareness Initiative, which is attached to C3I Division. His current interests lie in information fusion, reasoning under uncertainty, formal logics, conceptual structures and knowledge representation

---

Damian O'Dea

Command, Control, Communications & Intelligence Division

Damian O'Dea is a professional officer in Information Exploitation Discipline. His current work is involved in developing interoperability solutions for Command and Control Systems in coalition environments, developing information (data and semantic) models for preparedness reporting, and virtualisation of 3D command environments. He has worked in the past on geospatial information systems, middleware software solutions for defence applications, and distributed databases. He holds an Honours Degree in Computer and Information Science from the University of South Australia and is studying for a Master of Sciences (Defence).

Contents

1. Introduction

1.1 Data Fusion

In [1] and [2], Lambert defines data fusion broadly as

> …the process of utilising one or more data sources over time to assemble a representation of aspects of interest in an environment.

The traditional roots of the data fusion community are in sensor fusion, where the "data sources" are established sensors and the "aspects of interest in the environment" are moving objects, each typically represented by a set of state vectors. The broader definitions reflect an increasing emphasis toward generalising sensor fusion into so called higher-level fusion, in which "the aspects of interest in the environment" are not restricted to objects, but include biographic, economic, social, transport and telecommunications, geographic, military, political and technical information.

The Joint Directors of Laboratories (JDL) model was proposed in the late 1980s ([3]), with various revisions of it (e.g. [4], [5], [6], [7]) serving as the dominant model for data fusion. Figure 1 illustrates a variant of its revised form ([5]).

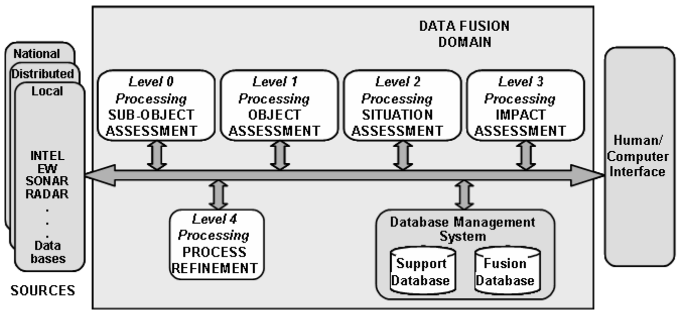

Figure 1: A revised JDL model of data fusion

The elements of the model are as follows:

- The source data provides representations of the world in numeric, graphic or symbolic form and can include surveillance, intelligence, public and other information.
- The sub-object assessments provide representations of detections of objects in the world, typically through numeric signal and/or image processing.
- The object assessments provide representations of objects in the world, typically through numerically based tracking and sensor fusion processing.
- The situation assessments provide representations of relations of interest between objects of interest in the world, typically through symbolic with some numeric processing, where the relations of interest can vary widely from concrete geospatial relations through to abstract political relations.

- The impact assessments provide representations of the consequences of interest from the representations of relations of interest between objects of interest in the world, typically through symbolic and some numeric processing, and involving threat assessments, course of action assessments and the like.
- The process refinements involve dynamic adaptations to sub-object, object, situation and impact processing, while also considering dynamic adaptations to sensor control.
- Databases are required to store the various representations.
- The system also needs to be able to interface with its human users.

1.2 Semantic Challenge

When contemplating machine based situation assessments, one confronts the question: "What symbols should be used and how do those symbols acquire meaning?" - termed "the Semantic Challenge" for Information Fusion by Lambert ([2]). The fusion system requires a means of representing the domain of interest in a meaningful way, and to do so, [2] notes that logics provide a means of imparting meaning to machines. A Situation Awareness by Inference and Logic (SAIL) pilot study has been established between the Defence Science and Technology Organisation (DSTO) and the National ICT Australia (NICTA) to investigate how logics can be employed to meet the Semantic Challenge. To provide a concrete approach, the SAIL programme has targeted a reduced version of the NATO North Atlantis scenario. Even in its reduced form, the scenario involves:

- nations and conflicts;
- physical geography;
- moving objects;
- military equipment with certain capabilities;
- civilian maritime and air traffic;
- masked intents – a military "chess game".

This document is primarily focussed on the conceptual aspects of the Semantic Challenge, *id est*, "What symbols should be used …". Within the SAIL programme, this relates to the Conceptual work package. Figure 2 illustrates the work packages comprising the SAIL programme and their dependencies.

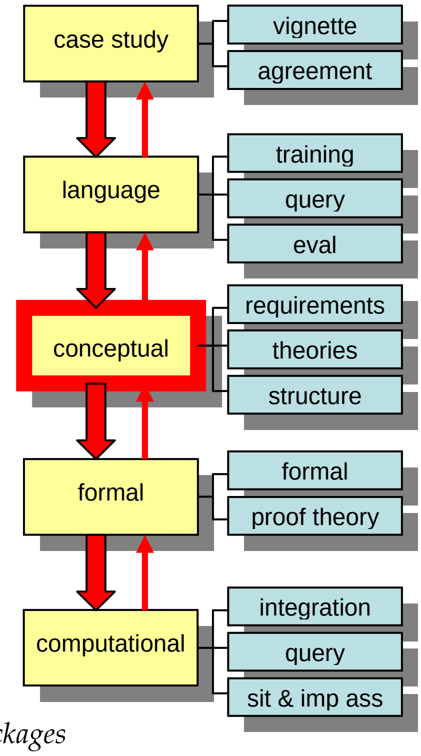

Figure 2: The SAIL Work Packages

Within the SAIL programme, milestone 2.1 states:

- Milestone 2.1: A joint document proposing the conceptual constructs required to express the selected aspects of the scenario. [31-Aug-07]

The DSTO contribution to this milestone was a document overview of conceptual frameworks (STE, NSM, CLCE, DOLCE, OpenCYC, Mephisto, SUMO, BFO, JC3IEDM) and a document outlining concepts of the Mephisto framework. The material of the first of these two documents is presented here in the form of a DSTO report.

1.3 Conceptual Design Choices

In choosing a conceptual framework for the SAIL programme, there are a number of design choices to consider. Some are listed below:

1. analytic vs. synthetic
2. descriptive vs. prescriptive
3. multiplicative vs. reductionism
4. formal vs. informal
5. realism vs. nominalism
6. endurantism vs. perdurantism
7. quantitative vs. qualitative
8. functionalism vs. non-functionalism
9. psychological vs. non-psychological
10. social vs. non-social

The *analytic* versus *synthetic* distinction concerns whether the framework is oriented toward information that can be determined independently from access to the world (analytic), or

whether it (also) includes information that can only be determined through access to the world (synthetic). For example, "all bachelors are unmarried males" is an analytic proposition while "Dale is wearing a white shirt" is a synthetic proposition. Analytic information is more oriented toward identifying the meaning of terms. Synthetic information is more oriented toward knowledge about the world.

The *descriptive* versus *prescriptive* distinction concerns whether the conceptual framework attempts to recover a commonsense, natural language based account of the world (descriptive), or whether it attempts to prescribe a conceptualisation for the world based on a philosophical or scientific framework (prescriptive). The information in the nightly television news services is descriptive. Quantum physics is prescriptive. A descriptive framework for the Conceptual module in Figure 2 will interface more easily to the Language module in Figure 2. A prescriptive framework for the Conceptual model in Figure 2 is likely to interface more easily to the Formal module in Figure 2.

The *multiplicative* versus *reductionism* distinction concerns whether the conceptual framework accepts a multitude of concepts as primitive (multiplicative), or whether it attempts to identify a small number of primitive concepts that are used to define concepts of interest (reductionism). High school accounts of life forms are usually multiplicative, with separate conceptualisation provided by Biology, Physics and Chemistry. The Peano axioms provide a reductionist conception of arithmetic based on the two primitives "zero" and "successor".

The *formal* versus *informal* distinction concerns whether the conceptual outlook is expressible in a formal language associated with a formal logic (formal), or not (informal). High school Biology concepts of life forms are typically informal while Peano's axioms are most commonly expressed in first order logic. Formal logics provide a means of resolving the remaining part of the semantic challenge "… and how do those symbols acquire meaning?". A formal conceptualisation provides an easier interface to the Formal work package in Figure 2. An informal conceptualisation may provide an easier interface to the Language work package in Figure 2.

The *realism* versus *nominalism* distinction concerns whether universals, like red and bigger, should be admitted as things in the world (realism) or whether the things of the world should be confined to particulars, or instances of properties and relations, only (nominalism). The nominalist typically confines the things of the world to pieces of matter and predicates over these. The realist typically allows abstractions like propositions, sets, possible worlds, or situations to also qualify as things that might be predicated.

The *endurantism* versus *perdurantism* distinction concerns whether an identity can exist at different times as an enduring object (endurantism), or whether an identity is formed from different things at different times as an assembly of different temporal parts (perdurantism). The endurantist accepts a world of enduring objects and confronts the problem of how an enduring object can change and yet still be the same (identity) object. The endurantist understands change in terms of things. The perdurantist accepts a world of processes and must confront the problem of how temporally different things can belong to the same (identity) thing. The perdurantist understands things in terms of change.

The *quantitative* versus *qualitative* distinction concerns whether the world can be understood quantitatively through various measures (quantitative) or qualitatively through various predicates (qualitative). These are not mutually exclusive, as the one conceptualisation can combine both. Quantitative concepts include measures such as temperatures and volumes, probabilities and geographical information system values. Qualitative concepts allow for abstract reasoning when concrete measurable details are unavailable or partially available. They include abstract conceptions, for example of space or time, and they include vague notions like "very" or "many".

The *functionalism* versus *non-functionalism* distinction concerns whether or not the conceptualisation is able to express functional accounts of entities in the world. This involves concepts that can describe functional parts of a system and the roles performed by those functional parts.

The *psychological* versus *non-psychological* distinction concerns whether or not the conceptualisation is able to express psychological accounts of entities in the world. This involves concepts that can describe mental states and how those mental states are formed.

The *social* versus *non-social* distinction concerns whether or not the conceptualisation is able to express social interactions between psychological entities in the world. This involves concepts that can describe various social constructs and how they arise.

The aforementioned alternatives have very real consequences for what can and cannot be achieved with a given conceptual framework and so provides an excellent basis for classifying conceptual approaches.

1.4 Existing Conceptual Frameworks

There are a number of existing conceptual frameworks that might be considered to address the conceptual requirements of the semantic challenge. These include:

1. AeroSpace and Defence Simplified Technical English;
2. Natural Semantic Metalanguage;
3. Descriptive Ontology for Linguistic and Cognitive Engineering;
4. OpenCyc;
5. Mephisto;
6. Suggested Upper Merged Ontology;
7. Basic Formal Ontology;
8. Object-Centred High-Level Reference Ontology; and
9. Joint C3 Information Exchange Data Model.

In the following sections, each of these is briefly discussed and classified with respect to the conceptual design choices noted in section 1.3.

2. Controlled Natural Languages

> Controlled Natural Languages are subsets of natural languages whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity. Traditionally, controlled languages fall into two major categories: those that improve readability for human readers, particularly non-native speakers, and those that improve computational processing of the text ([8]).

The diverse motivations driving the formation of a Controlled Natural Language (CNL) impact, to some degree, on its conceptual model and practical nature. Three CNLs are discussed here with respect to the conceptual design choices (outlined in §1.3) and additional assessment criteria relevant to the semantic challenge of data fusion, specifically, for Situation Awareness by Inference and Logic (SAIL). These should not be taken as the only relevant criteria, but are used to assess each CNL's important, essential and extant/missing or deficient features.

Three CNLs are assessed according to particular criteria (see §2.1). These languages are AeroSpace and Defence Simplified Technical English (henceforth ASD-STE[1] or simply STE) ([9]), Common Logic Controlled English (henceforth CLCE) ([10]) and Natural Semantic Metalanguage (henceforth NSM) ([11]). The three have been selected in part because, in some respects, they each typify diverse approaches to the problem.[2] In general terms, we identify a natural–formal scale in which 'natural' represents a CNL that is closest to natural human language and 'formal' represents a CNL that is closest to a machine language. We can categorise the three CNLs according to the assessment criteria below. ASD-STE is perhaps closest to the natural human 'informal' pole; CLCE is closest to the machine language formal end, and NSM is placed somewhere between both, as represented in Figure 3.

'natural' human language ⟷ 'formal' machine language

ASD-STE NSM CLCE

Figure 3: Categorisation of three Controlled Natural Languages according to natural-formal scale

For instance, ASD-STE incorporates and selects word senses from those extant in natural language. In contrast, NSM is conceptually grounded in postulated universal primitives and prescribes their formal implementation. It claims to identify the universal semantic invariant of a relatively small set of semantic primes. CLCE focuses on the ability to represent formal logic in a natural-like fashion which approximates natural language while maintaining computational 'well-formedness'.

---

[1] In Section 2.2 a specific version is discussed and abbreviated to ASD-STE100. We refer to it in this section in general terms and thus omit 100.

[2] An additional appeal of ASD-STE is that the domain may be relevant to that of SAIL.

2.1 Assessment criteria

Clearly, the criteria for assessing the appropriateness of a particular CNL needs to be based, to a large extent, on the purpose to which it is set. At this point the goal is an adequate conceptualisation for the representation of knowledge with the practical caveat of bidirectional mapping between a formal and natural language. In other words, our ideal CNL needs both to be conceptually sound and to function as an interlanguage between a human and an artificial computational entity.

The underlying conceptualisation of each of the CNLs discussed here are direct reflections of the goals of the tasks which drove their individual formulations. This has resulted in two considerably different conceptual and practical models. Moreover, each CNL addresses—to varying degrees—issues of expressivity, coverage, robustness, human comprehensibility, computational adequacy and formalness. These criteria are each briefly outlined.

2.1.1 Conceptualisation

Conceptualisation refers here to the conceptual drivers that develop from underlying beliefs upon which each CNL is built. For instance, we can observe that a driver of the Natural Semantic Metalanguage (NSM) is an attempt at fundamental explanatory adequacy with respect to human language and cognition, viz innate, primitive (irreducible) concepts, and a psychologically feasible model of language acquisition ([11] p. 16-19). For AeroSpace and Defence's Standard Technical English (ASD-STE) ([9]) a conceptual driver is the practical desire for standardisation of language use. For Common Logic Controlled English (CLCE) ([10]) a driver is the degree to which it can be made maximally readable without loosing the tight mapping to first-order logic.

Unifying each of these approaches is the concept of a controlled language whose expressive power resembles 'natural' human language but which also strives for something 'unnatural' to human language; isomorphism of form and meaning, that is the disambiguation of meaning—a stripping away of the possibility of multifarious interpretations through prescribed one-to-one association between meaning and form.

2.1.2 Expressivity

This pertains to the degree to which the language can be used in a diverse range of contexts and registers. Can it, for instance, express hypothetical or modal propositions or only declarative facts in a restricted tense? Each CNL is formulated to attain a degree of 'naturalness' and ease of use for humans to compose. For one CNL, the grammar emerges on-the-fly, through a kind of negative elimination of proscribed terms (ASD-STE), while for others (NSM & CLCE) rigid construction rules are stipulated. Further, CLCE is limited in expressivity to propositions of first-order logic; e.g., present tense verbs and singular nouns only.

2.1.3 Coverage

This pertains to the object domain; aerospace, defence, "primitive" concepts and speech acts or propositional logic. To some extent, the specific domain conditions both the range of terms and particular constrained interpretations thereof. Thus, ASD-STE is more elaborated in the aerospace and defence domain than the other CNLs, whereas NSM has greater coverage with respect to speech acts.

CLCE and ASD-STE are distinguished from NSM in terms of the open lexicon, whereas NSM is restricted to a closed set. However, NSM makes the strong claim that all and every concept can be expressed with a closed set of universal primes ([11] p.15). It should be noted that coverage and expressivity are closely dependent. For instance, the field of aerospace may require particularly elaborate coverage of geospatial terms and the ability to express them appropriately by the grammar. Further, although a CNL may be elaborated in some domain with specific technical terms, a base-line of natural language vocabulary will be required by any CNL for the use of basic words and structures of everyday language.

2.1.4 Robustness

This pertains to the degree of completeness, internal consistency and interconnectedness of the grammar; i.e., it should not suddenly fail to interpret alternate formulations or paraphrases. In other words, it should be able to handle and interpret multiple variations of any (or at least most) NL formulation of the communicative intent.

2.1.5 Human comprehensibility

This pertains to the degree to which a formulation in a specific CNL is easy to interpret for humans in terms of their own natural language intuitions.

2.1.6 Computational adequacy

This pertains to the capacity to map to a computation mechanism. CLCE is the only CNL discussed here which is specifically designed for computational adequacy. It is claimed ([10] p. 1) to be fully 'compilable' to an implementation language. As far as we know, no attempts have been made to make either NSM or ASD-STE computationally adequate.

2.1.7 Formalness

This pertains to the degree to which the CNL adheres not only to strict conventions, but to constitute a system that is rigorous and unambiguous. In addition, we need to differentiate a well-formed formal specification language from one that is formal but not (computationally) axiomatised. In these respects, ASD-STE is not a formal language; NSM is formal and semantically and syntactically rigorous but not a specification; whereas CLCE claims to be a formal specification language which will produce well-formed formulae.

We now turn to discuss some specific features of each of the CNLs in slightly more detail.

2.2 AeroSpace and Defence Simplified Technical English (100)

2.2.1 Introduction

The AeroSpace and Defence Simplified Technical English (ASD-STE100) is an 'international specification for the preparation of maintenance documentation in a controlled language' ([9] p.i.)

ASD-STE100 is comprised of a set of Writing Rules and a Dictionary of controlled vocabulary. This syntax and lexicon both prescribe and proscribe the use of particular terms and their use. For example, ASD-STE100 states that:

> When there are several words in English for a certain thing or action (synonyms), this Specification gives one of these synonyms to the exclusion of the others (whenever possible, "one word – one meaning"). For example, "start" was chosen instead of "begin", "commence", "initiate", or "originate". … When there are several possible definitions of a word in English, in general the Specification give one of these definitions to the exclusion of the others… For example, "to fall" has the definition of "to move down by the force of gravity", not "decrease" ([9] p. i).

There are a number of problems with these statements, briefly discussed here.

2.2.1.1 Unmotivated proscription of terms

The specification document [9] does not stipulate the basis for the choice of one term over and above another. For instance, according to what principle is the noun 'place' proscribed in preference for the noun 'area'? Without this information it is difficult to motivate principled lexicographic treatments, which have the potential to be translated into automated computational algorithms.

2.2.1.2 "One word – one meaning" principle

As is typical of all languages, words carry multiple meanings. In spite of the stated preference for the principle of monosemy, 'one word-one meaning', many of the ASD-STE100 prescribed words are polysemous. That is, ASD-STE100 prescribes monosemy, but in fact sanctions multiple senses. For instance, the specification gives two distinct senses for 'arm' as a verb: 1. 'to install armaments' and 2. 'to prepare for automatic operation'. Further, it gives two undifferentiated senses for the noun 'area': 'a specified surface or location'. Leaving aside the issues of whether these are adequate definitions[3], there is no indication of the relative importance of each sense of 'arm' and in which semantic or syntactic contexts they are to be used. Further, the formulation of the assigned meaning of 'area' is itself ambiguous: 'a specified surface or location'. As the scope of the modifier is left unspecified, this could mean: 'a specified surface or a specified location' in which the modifier 'specified' has scope over both nouns or 'a specified surface' or 'location' in which it only has scope over the adjacent noun.

---

[3] It could be argued that the definition of sense 1. 'to install armaments' will turn out to be circular because the term 'armaments' contains the word 'arm' and so this definition is defined in terms of itself.

Two repercussions of these failures to adhere to the 'one word–one meaning' policy are that (i) ambiguity and vagueness will be left to the human, who uses the language to encoded information, to identify and resolve (and, where this is not done, the receiving person, the reader, may be unable to resolve ambiguities); and (ii) a component of the task of mapping to an axiomatic computational language will need to consider how to account for, or at the very least be aware of, multiple senses of lexemes.

2.2.2 Analysis

Assessed from the conceptual framework outlined in §1.3, ASD-STE100 attests the following tendencies:

1. **analytic** vs. synthetic
2. **descriptive** vs. prescriptive
3. **multiplicative** vs. reductionism
4. formal vs. **informal**
5. realism vs. **nominalism**
6. **endurantism** vs. perdurantism
7. **quantitative** vs. qualitative
8. functionalism vs. **non-functionalism**
9. psychological vs. **non-psychological**
10. social vs. **non-social**

2.2.3 Standard Technical English 100: lexicon and grammar

We do not analyse here ASD-STE in detail, but give some brief comments on the lexicon and grammar in order to give a comparative flavour for the language and as a preface to the discussion of a selection of verbs, §2.2.4.

ASD-STE100 prescribes a set of approved 'keywords'. This constitutes the basic lexicon (or word-list) for use. However, the specification states that additional words may be used where necessary, but does not describe what would constitute a necessary condition.

Parts of the grammar guide for using ASD-STE100 are rather general and thus somewhat unclear as to its intention as a recommendation for the copywriter (the assumed intent) or a description of an already written product, such as a maintenance manual:

> When it is possible, the assigned meanings are written in Simplified Technical English. If it is necessary to use words that are not approved in STE, these words are shown in quotation marks ("…"). If a Technical Name or a Technical Verb is used in an assigned meaning, this word is identified as (TN) or (TV). ([9] p. 2-0-5).

The 'Dictionary' lists each keyword (with approved forms) together with its part of speech in parenthesis. Apart from technical terms, it stipulates the use of only approved words and only in the form and manner determined by its function when used in the part of speech shown.

There are eight parts of speech identified: verb (v); noun (n); adjective (adj); adverb (adv); article (art); preposition (pre); pronoun (pn); conjunction (con). There are no sub-class distinctions, other than pronoun. This is not given as a sub-class of noun, but as a separate part of speech. The motivation for this is unclear. Further information on word classes is that '[t]he last four parts of speech are "function words". Functions words connect and relate ideas in a sentence' ([9] p. 2-0-4). This description is not instructive.

The use of the passive voice is proscribed in 'procedural writing and as much as possible in descriptive writing' ([9] p. 1-3-3). No explanation for this is given. It is not necessarily the case that the use of the passive voice obfuscates the communicative intent. Rather in some situations its use would be considered appropriate.

2.2.4 Prescribed verbs

We now present in tabular form a selection of verbs from the ASD-STE100 dictionary, Table 1. We focus on verbs, as these are the most complex; conceptually, formally, and computationally.

Column one presents the keyword with its prescribed part of speech followed by the permissible forms.[4]

Column two presents the ASD-STE100 'Assigned Meaning' (AM).[5] This is not a full explication of the term's meaning but a one-line guide probably intended to trigger recognition by the human user of the appropriate prescribed sense. Although ASD-STE is intended for use by both native- and non-native-speakers of English, the AM is perhaps too terse to provide enough information for the non-native-speaker to grasp the full conditions on its application.

Column three presents the 'Approved Example'. This is an example of the keyword in use and goes some way to filling in the information missing from the AM. These first three columns repeat verbatim the information in ASD-STE100.

Column four contains the closest sense to the AM from the Oxford Advanced Learner's Dictionary (OALD) ([12]). This is shown with the acronym oald_#, where # is the number of the closest (partially) equivalent sense from the set enumerated in the OALD.

Column five contains comments on problems concerning interpretation of meaning.

---

[4] In the original document this column also contains proscribed words with part of speech.

[5] In the original document this column also contains the prescribed keywords in cases where column contains proscribed words.

Column six contains somewhat terse analyses of predicates based on the AM and Approved Example, viz transitivity, thematic roles of arguments and grammatical relations.

Transitivity refers to the number of arguments selected by the verb. Here vt = transitive verb (2 arguments), vc = copular verb (links arguments).

Thematic roles 'connect entities to events (more particularly, arguments to predicates)' ([13] p. 197).[6] There is no universally identified fixed set of thematic roles. Twelve are exemplified here:

- **Agent**: deliberately performs the action (e.g., **Bill** ate his soup quietly)
- **Experiencer**: receives sensory or emotional input (e.g., The smell of lilies filled **Jennifer's** nostrils).
- **Theme**: undergoes the action but does not change its state (e.g., Bill kissed **Mary**). (Sometimes used interchangeably with patient)
- **Patient**: undergoes the action and has its state changed (e.g., The falling rocks crushed **the car**) (Sometimes used interchangeably with theme)
- **Instrument**: used to carry out the action (e.g., Jamie cut the ribbon **with a pair of scissors**).
- **Natural Cause**: mindlessly performs the action (e.g., **An avalanche** destroyed the ancient temple).
- **Location**: where the action occurs (e.g., Johnny and Linda played carelessly **in the park**).
- **Goal**: what the action is directed towards (e.g., The caravan continued on **toward the distant oasis**).
- **Recipient**: a special kind of goal associated with verbs expressing a change in ownership, possession. (e.g., I sent **John** the letter)
- **Source**: where the action originated (e.g., The rocket was launched **from Central Command**).
- **Time**: the time at which the action occurs (e.g., The rocket was launched **yesterday**)
- **Beneficiary**: the entity for whose benefit the action occurs (e.g., I baked **Reggie** a cake)

Grammatical relations refers to the syntactic relationship of arguments to the predicate, viz, subject (S), dummy-subject (dS), (direct or primary) object (O1), secondary object (O2), indirect object (IO), oblique object (OO).

It is recognised that the analyses presented in this column—while crucial for a comprehensive treatment of requirements—are too concise to be functionally useful. They are given here as indicative of the necessary analyses. A substantial treatment is proposed for a forthcoming document as part of the Language work package.

---

[6] These are also labelled as: thematic roles, semantic cases, thematic relations, θ (theta) roles (see [13] p. 197 and references therein). The list given here varies somewhat to that in Frawley [13] Ch 5. No theoretical commitment is made here to the current explanatory list.

Table 1: Selection of 12 verbs from the ASD-STE100 dictionary with comparisons, comments and analyses

Verbatim from ASD-STE100			Comparisons, comments and concise analyses		
Keyword (part of speech), other forms	**ASD-STE100 Assigned Meaning**	**ADS-STE100 Approved Example**	**Closest OALD sense**	**Comments (AM=assigned meaning)**	**Transitivity: Thematic roles of AM: Grammatical relations:**
ABSORB (v), ABSORBS, ABSORBED, ABSORBED	1. To "take" up or into	Absorb the fluid with a clean cloth.	oald_1 liquid/gas: to take in a liquid, gas or other substance from the surface or space around	Unclear explanation in AM of 'take up' re: 'absorb'. Problem of difference in sense with 2.	vt agent, patient, instrument S,O1,OO
ABSORB as above	2. To decrease the effect of	The shock mount absorbs the vibration.	oald_6 shock/impact: to reduce the effect of a blow, hit, etc.	Problem of difference in sense with 1. Polysemous or vague sense?	vt receptor, force S,O1
ACCEPT (v), ACCEPTS, ACCEPTED, ACCEPTED	To make a decision that sth is satisfactory.	Accept the reality if it is serviceable.	oald_2 receive as suitable: to receive sth as suitable or good enough	AM unclear, i.e., AM & AE do not equate	vt decider, theme S,O1
ADAPT (v), ADAPTS, ADDAPTED, ADDAPTED	To change or adjust to that which is necessary	Adapt the pressure connection to the pitot head	oald_1 adapt sth (for sth): to change sth in order to make it suitable for a new use or situation	AM unclear how it is diff from 'adjust', but see thematic roles.	vt agent, patient S,O1
ADD (v), ADDS, ADDED, ADDED	To increase the number	Add 5 milliliters of hardener to the compound	oald_1 add sth (to sth): to put sth together with sth else so as to increase the size, number, amount, etc.	AM is specialisation to 'number' only vs. oald_1 'size', 'amount', …	vt agent, medium, medium S,O1,O2
ADJUST (v), ADJUSTS, ADJUSTED, ADJUSTED	To put to a specified position or "state"	Adjust the controls carefully	oald_1 adjust sth (to sth) to change sth slightly to make it more suitable for a new set of conditions or to make it work better	AM does not refer to degree of change: e.g., oald_1	vt agent, feature S,O1

AGREE (v), AGREES, AGREED, AGREED	To be "consistent" with	The indications must agree with the values in the table	oald_6: to be the same as sth	AM highly specialised usage, rather unintuitive. Focal sense has cognitive agents in thematic roles.	vt measurement, measurement S,O1,O2
ALIGN (v), ALIGNS, ALIGNED, ALIGNED	To put into line (TN) with	Align the flange holes with the studs	oald_1: to arrange sth in the correct position, or to be in the correct position, in relation to sth else, especially in a straight line	Sense appears more appropriate for desired sense of AGREE	vt agent, theme, theme S,O1,O2
APPLY (v), APPLIES, APPLIED, APPLIED	1. To put on	Apply a force of 100N on the end of the lever	oald_6 press hard: to press on sth hard with your hand, foot, etc. to make sth work or have an effect on sth	Specialised sense of oald_6	vt agent, patient, force S,O1,O2
APPLY as above	2. To "spread" on	Apply the sealing compound	oald_3 paint/cream: to put or spread sth such as paint, cream, etc. onto a surface	Problem of difference in sense with 1	vt agent, patient, goal S,O1,(O2)
ARM (v), ARMS, ARMED, ARMED	1. To install armaments	Arm the aircraft	oald_1 to provide weapons for yourself/sb in order to fight a battle or a war	AM is a specialisation of the oald_1 sense	vt agent, patient S,O1
ARM as above	2. To prepare for automatic operation	Set the switch to "close" to arm the circuit	oald_2 to make a bomb, etc. ready to explode	AM has specialised additional component of 'automatic operation'	vt agent, patient S,O1
ASSEMBLE	To attach items together	Assemble the valve as follows	oald_2 to fit together all the separate parts of sth, for example a piece of furniture	AM is a broadening and has lost the requirement of 'all … parts'	vt agent, patient S,O1
BALANCE	To make equal	Balance the electrical loads on each generator	oald_v_2 to be equal in value, amount, etc. to sth else that has the opposite effect	AM is a specialised sense with loss of 'opposite effect', see thematic roles	vt agent, patient S,O1

BE (v), IS, WAS (also ARE, WERE), NOTE: No other forms of this verb	1. To occur, “exist”	If there is corrosion on the pump vanes, replace the pump	oald_1 to exist; to be present	Usually has presentative function, but not described as such. Problem of difference in sense with 2.	vc Ø, entity dS,O1
BE as above	2. To have a property, to be equal to	These solutions are dangerous	oald_14 to be equal to	AM does not differentiate between two senses: have property & be equal. Can be analysed as class-membership.	vc member, class S,O1

This cursory survey of 12 verb lexemes shows numerous problems. There are inadequate definitions of the Assigned Meaning, e.g., 'absorb'. There are discrepancies between the Assigned Meaning and the actual sense when used in the Approved Example, e.g., 'accept'. There are both explicitly and implicitly sanctioned polysemous senses, e.g., 'arm', 'be' and 'be$_2$' respectively. There are unclear choices of particular senses as the Assigned Meaning which clearly deviates from a native-speaker's natural language intuition, e.g., 'agree'. There are seemingly unmotivated specialisations of particular senses, e.g., 'agree'. There are unmotivated broadenings of senses, e.g., 'assemble'. The Assigned Meaning is mostly a reflex of a specific lexical sense and does not include the joint role of lexical and constructions contributions to meaning, e.g. 'be' with a dummy subject as a presentative construction.

In sum, ASD-STE does indeed count as a controlled language in the sense that it specifies a sub-set of natural language lexicon and grammar. However, we have described numerous aspects where it breaks its own rules and the result is a loss of rigour. Without considerable effort, this would make the language very difficult to map to a formal language.

2.3 Common Logic Controlled English

2.3.1 Introduction

Common Logic Controlled English (CLCE) is a formal language with an English-like natural syntax. Anyone who can read ordinary English can read sentences in CLCE with little or no training. This is potentially very useful for the SAIL requirements. Writing CLCE, however, requires practice in learning to stay within its syntactic and semantic limitations. Formally, CLCE supports full first-order logic with equality supplemented with an ontology for sets, sequences, and integers. The fundamental semantic limitation of CLCE is that the meaning of every CLCE sentence is defined by its translation to FOL; none of the flexibility of ordinary English and none of its metaphorical or metonymic extensions are supported. The primary syntactic restrictions are the use of present tense verbs and singular nouns, variables instead of pronouns, and only a small subset of the many syntactic options permitted in English. Despite these limitations, CLCE can express the kind of English used in software specifications, textbooks of mathematics, and the definitions and axioms of formal ontology ([10]).

1. **analytic** vs. synthetic
2. descriptive vs. **prescriptive**
3. multiplicative vs. **reductionism**
4. **formal** vs. informal
5. realism vs. **nominalism**
6. endurantism vs. **perdurantism**
7. **quantitative** vs. qualitative
8. functionalism vs. **non-functionalism**
9. psychological vs. **non-psychological**
10. social vs. **non-social**

CLCE is bound by some fundamental semantic limitations whereby the meaning of every CLCE sentence is defined by its translatability to first-order logic (FOL). That is, it claims not to support any extensions of metaphor or metonymy ([10] p. 1).

2.3.2 Bidirectionality

CLCE claims to be a bidirectional language which permits translation from CLCE and FOL. However, Sowa ([10] p. 1-2) states that if the FOL statement has not been derived from an original CLCE formulation, then equivalence of CLCE to FOL can only be ensured if appropriate mappings between CLCE and FOL terms are manually checked.

Logical equivalence between CLCE and FOL is claimed to hold, even when a reverse translation results in a CLCE statement that is not identical to the input.

The ability to check the equivalence of a paraphrase with an original input through back-translation should be recognised as a potentially useful means of system verification and reliability. The issue of the human operator's overall 'trust' of the system should not be underestimated. The ability to present the system with a proposition and require it to return an essentially synonymous paraphrase would be one way to build this relationship between user and system.

2.3.3 Expressivity

The tight grammatical restrictions of CLCE significantly reduce its expressivity. We briefly survey key issues here.

2.3.3.1 Tense

Verb forms are constrained to the present tense. This means that, unless temporal relations were specifically stipulated as other than the default interpretation, events are expressed as holding at the time of utterance. Clearly, this is insufficient for the requirements of SAIL where the differentiation of fine-grained temporal relations is of paramount relevance.

2.3.3.2 Number

Nouns of only singular number can be represented. Group nouns can be used to represent collections, such as 'pair' with two arguments or 'group' with an arbitrary list of arguments. The inability to represent number in the way of a natural language severely curtails the 'naturalness' of CLCE. Moreover, it is at times preferred to be vague with respect to the number of an expression, such as 'some vehicles are approaching', where it is known that the set is more than one or two but the precise number is not known.

2.3.3.3 Pronouns

Variables are used in place of pronouns. This is a way to overcome the problematic issue of phoric reference resolution with natural language terms such as pronominals and deictics.

2.3.3.4 Voice

Through the allocation of thematic roles (see §2.3.4 below) to argument positions in predicate structure, CLCE provides a means to handle both active and passive voice formulations of a proposition. This is useful because the ability to process voice alternations is predicted to be a likely requirement of the CNL component and would go some way to increasing its 'naturalness'.

2.3.4 Thematic roles

Many treatments of argument structure of natural language propose the mapping of syntactic argument slots with thematic roles. Sowa ([14] p. 506) adopts this general approach in an idiosyncratic way. For instance, rather than attributing a participant with a particular role at a given time, Sowa sets up an ontology of Participant, to which thematic roles are classified as subtypes. This is a somewhat difficult approach to pursue, as it fails to address the well attested possibility of a single participant holding thematic multiple roles ([15] p. 59-70).

2.3.5 Output

A potential strength of Sowa's approach is its quasi-natural language form. It has been developed in concert with the Conceptual Graph Interchange Form (CGIF). The following are some examples of CLCE compared with translations to Predicate Calculus (PC) and CGIF:

(1) a. CLCE: **`Every cat is on a mat.`**
    b. PC: **`(∀x:Cat) (∃y:Mat)On(x,y)`**
    c. CGIF: **`(On [Cat: @every] [Mat])`**

(2) a. CLCE: **`Some person is between a rock and a hard place.`**
    b. PC: **`(∃x:Person)(∃y:Rock)(∃z:Place)`**
    `(Between(x,y,z)&Hard(z))`
    c. CGIF: **`(Between [Person] [Rock] [Place *z]) (Hard ?z)`**

2.3.6 Elimination of ambiguity

The resolution of ambiguity is an essential component of the CNL. CLCE has provision for restricting potential ambiguity in the following ways:

(i) CLCE claims to avoid multiple words senses. Thus, each word is designated a single sense and this is encapsulated in the syntactic formulation of a single relation.

(ii) CLCE limits the type of prepositional phrases that can combine with each major part of speech. This is designed to constrain possible syntactic structure to control semantics. Only a single preposition **`of`** can be attached to a noun. Other prepositions can only attach to verbs.

(iii) Quantifiers have limited scope. Universal quantifiers are limited to the subject of a clause or a prefix before a clause. Existential quantifiers can occur elsewhere. This allows the syntax to determine their scope.

(iv) Pronouns are replaced by explicit variables. This addresses the problem for anaphoric resolution in natural language.

(v) Nominal modifiers, other than adjectivals such as colour, must be pre-combined with nouns to form a single modified predicate as in **`hard_disk_drive`**. This avoids extra mechanisms for processing them. However, there is need for more detailed description of the precise range of modifiers and how they are each handled.

(vi) Sentential and clausal embedding must be explicitly represented through the use of bracketing.

(vii) Natural language features that are currently beyond the ability of FOL are not permitted in CLCE. These include plural noun phrases, verb tenses, modality and an open-ended number of context-dependent questions.

2.3.7 The Grammar: lexicon and syntax

CLCE has provision for three major utterance types: declarative, interrogative and imperative. These three types are a minimal requirement for any CNL system. CLCE also has special 'extended sentences', separated by a semicolon. The function of this sentence type is to allow the scope of variables to extend beyond a single sentence.

The lexicon of CLCE is split between *reserved* and *declared* words. Reserved words are restricted in meaning to a single sense or a limited set determined by syntactic context. There are ten kinds of restricted words shown in

1. **Boolean operators:** not, and, either, or, neither, nor, if, then.
2. **Quantifiers:** a, an, some, something, someone, every, everything, everyone, no, nothing, no one.
3. **Special verbs:** is, has, have, does.
4. **Interrogatives:** who, what, when, where, which.
5. **Relative pronoun:** that.
6. **Definite article:** the.
7. **List connector:** and.
8. **Special lists:** none, others, nothing else, no one else.
9. **Argument markers:** of, than, as.
10. **Special phrases:** there is, such that, only if, if and only if, it is false that, is it true that.

Declared words can represent any of the following: names, nouns, verbs, adjectives, adverbs, or prepositions. For a word to be used in CLCE it must be mapped to a FOL relation by declaration.

The grammar component uses defined lexical categories of the following types:

Table 2: Lexical categories of CLCE grammar component

<table>
<tr><th>Category</th><th>Description</th><th>Example</th></tr>
<tr><td>Adjective</td><td>Any word that has been declared as an adjective.</td><td><code>Adjective = Word</code></td></tr>
<tr><td>Character string</td><td>Zero or more characters delimited by double quotes. Any double quote in a character string must be doubled.</td><td><code>CharString = """", {Character - """" | """"""}, """"</code></td></tr>
<tr><td>Comparative</td><td>Any word that has been declared as a comparative.</td><td><code>Comparative = Word</code></td></tr>
<tr><td>Ellipsis</td><td>A string of three periods.</td><td><code>Ellipsis = "..."</code></td></tr>
<tr><td>Exponent</td><td>The letter E in upper or lower case followed by an optional sign and an unsigned integer.</td><td><code>Exponent = ("e" | "E"), ["+" | "-"], UnsignedInt</code></td></tr>
<tr><td>Floating-point number</td><td>An optional sign followed by one of three options: (1) a decimal point, an unsigned integer, and an optional exponent (2) an unsigned integer, a decimal point, an optional unsigned integer, and an optional exponent or (3) an unsigned integer and an exponent.</td><td><code>Floating = ["+" | "-"], (".", UnsignedInt, [Exponent] | UnsignedInt, (".", [UnsignedInt], [Exponent] | Exponent))</code></td></tr>
<tr><td>Functional noun</td><td>Any word that has been declared as a functional noun and with a relational pattern.</td><td><code>FunctionalNoun = Word</code></td></tr>
<tr><td>Integer</td><td>An optional sign followed by an unsigned integer.</td><td><code>Integer = ["+" | "-"], UnsignedInt</code></td></tr>
<tr><td>Name</td><td>Either a quoted name or any word that has been declared as a name.</td><td><code>Name = QuotedName | Word</code></td></tr>
<tr><td>Noun</td><td>Any word that has been declared as a preposition.</td><td><code>Noun = Word</code></td></tr>
<tr><td>Number</td><td>Either an integer or a floating-point number.</td><td><code>Number = Integer | Floating</code></td></tr>
<tr><td>Preposition</td><td>Any word that has been declared as a preposition.</td><td><code>Preposition = Word</code></td></tr>
<tr><td>Quoted name</td><td>Zero or more characters delimited by single quotes. Any single quote in a quoted name must be doubled.</td><td><code>QuotedName = "'", {Character - "'" | "''"}, "'"</code></td></tr>
<tr><td>Relational noun</td><td>Any word that has been declared as a noun with a relational pattern.</td><td><code>RelationalNoun = Word</code></td></tr>
<tr><td>Unsigned integer</td><td>A sequence of one or more digits.</td><td><code>UnsignedInt = Digit, {Digit}</code></td></tr>
<tr><td>Variable</td><td>One of the four letters <code>x</code>, <code>y</code>, <code>z</code>, or <code>w</code> in either upper or lower case followed by an optional string of digits.</td><td><code>Variable = ("x" | "X" | "y" | "Y" | "z" | "Z" | "w" | "W"), {Digit}</code></td></tr>
<tr><td>Verb</td><td>Any word that has been declared as a verb.</td><td><code>Verb = Word</code></td></tr>
<tr><td>Word</td><td>Any sequence of one or more letters, digits, or underscores that is not a variable or an integer.</td><td><code>Word = (((Letter | Digit | "_"), {Letter | Digit | "_"}) - Variable) - Integer</code></td></tr>
</table>

There are restrictions on how the lexicon can be combined. This is governed by syntactic categories. Possible syntactic categories are:

Table 3: Syntactic categories of CLCE

Category	Description	Example
Comparison	A comparative followed by the word than and a simple noun phrase.	`Comparison = Comparative, "than", SimpleNP;`
Complex declarative sentence	Either a conditional clause followed by ", then" and a parenthesised declarative sentence or a parenthesised declarative sentence followed by either "only" and a conditional or "if and only" and a conditional. Either a conditional clause followed by , then and a parenthesised declarative sentence or a parenthesised declarative sentence followed by either only and a conditional or if and only and a conditional.	`ComplexDS = Conditional, ", then", ParenDS | ParenDS, ( "only", Conditional | "if and only", Conditional);`
Compound declarative sentence	Two or more parenthesised declarative sentences connected by and or by or. If the connector is and, each sentence except the last must be followed by a comma, and the last sentence must be preceded by and. If the connector is or, the first sentence must be preceded by either, each sentence except the last must be followed by a comma, and the last sentence must be preceded by or.	`CompoundDS = ParenDS, "," {ParenDS, ","}, "and", ParenDS | "either", ParenDS, "," {ParenDS, ","}, "or" ParenDS;`
Conditional clause	The word if followed by a declarative sentence.	`Conditional = "if", Declarative;`
Declarative sentence	A sequence of zero or more logical prefixes, followed by a simple declarative, a complex declarative, or a compound declarative:	`Declarative = {LogPrefix}, (SimpleD | ComplexD | CompoundD);`
Existential noun phrase	Either the word something or someone followed by an optional variable; the word the followed by a functional noun, an optional variable, the word of, and a list; or one of the three words a, an, or some followed by a term.	`ExistentialNP = ("something" | "someone"), [Variable] | "the", FunctionalNoun, [Variable], "of", List | ("a" | "an" | "some"), Term;`
Extended sentence	A sequence of one or more sentences separated by semicolons and ending with a period or a question mark. There are three kinds of extended sentences: one or more declarative sentences ending with a period; one or more imperative sentences ending with a period; or zero or more declarative sentences followed by one or more interrogative sentences ending with a question mark.	`ExtendedSentence = Declarative, {";", Declarative} "."; | Imperative, {";", Imperative} "."; | {Declarative, ";"}, Interrogative, {";" Interrogative}, "?";`
Generalised name	Either a name, a variable, a number, or a character string.	`GeneralName = Name | Variable | Number | CharString;`
Interrogative sentence	An optional conditional clause and comma followed by either a WH-phrase and a verb phrase or the phrase is it true that and a declarative sentence.	`Interrogative = [Conditional, ","], ( WhPhrase, VerbP | "is it true that", Declarative);`

Category	Description	Example
List	Either one simple noun phrase, optionally followed by the word and and another simple noun phrase, or a parenthesised list of one or more simple noun phrases, each followed by a comma, and either an ellipsis or the word and followed by either the word others, the phrase nothing else, the phrase no one else, or a simple noun phrase.	`List = SimpleNP, ["and", SimpleNP] | "(", SimpleNP, ",", {SimpleNP, ","},( Ellipsis | "and", ("others" | "nothing else" | "no one else" | SimpleNP)), ")";`
Logical prefix	Either the word for followed by a universal noun phrase and a comma, the words there is followed by an existential noun phrase and the words such that, or the words it is false that.	`LogPrefix = "for", UniversalNP, "," | "there is", ExistentialNP, "such that" | "it is false that";`
Negated existential noun phrase	Either the word something or the phrase no one followed by an optional variable or the word no followed by either a term or a functional noun, an optional variable, the word of, and a list.	`NegExistentialNP = ("nothing" | "no one"), [Variable] | "no", (Term | FunctionalNoun, [Variable], "of", List);`
Noun Phrase	One of four kinds: existential, universal, negated existential, or referential.	`NounP = ExistentialNP | UniversalNP | NegExistentialNP | ReferentialNP;`
Parenthesised declarative sentence	A declarative sentence in which any compound or complex declarative sentence is enclosed in parentheses:	`ParenDS = {LogPrefix}, (SimpleDS | "(", ComplexDS, ")" | "(", CompoundDS, ")" );`
Postmodifier	Either a comparison or the word that followed by a verb phrase.	`Postmodifer = (Comparison | "that", VerbP);`
Predicate complement	An optional not followed by either an adjective, a simple noun phrase, a comparison, or a prepositional phrase.	`PredComp = ["not"], (Adjective | SimpleNP | Comparison | PrepositionalP);`
Prepositional phrase	A preposition followed by a list.	`PrepositionalP = Preposition List;`
Referential noun phrase	Either a generalised name or the word the followed by either a noun, a relational noun, or a functional noun and an optional general name.	`ReferentialNP = GeneralName | "the", (Noun | RelationalNoun | FunctionalNoun),[GeneralName];`
Sentence	Either declarative, interrogative, or imperative.	`Sentence = Declarative | Interrogative | Imperative;`
Simple declarative sentence	Either a sequence of zero or more prepositional phrases, each followed by a comma (none of which may begin with the preposition for) followed by a noun phrase and a verb phrase; or the phrase there is followed by either an existential noun phrase or a negated existential noun phrase.	`SimpleDS = {PrepositionalP - ("for", SimpleNP), ","}, NounP, VerbP | "there is", (ExistentialNP | NegExistentialNP);`
Simple noun phrase	Either an existential noun phrase or a referential noun phrase.	`SimpleNP = ExistentialNP | ReferentialNP`

Category	Description	Example
Simple verb phrase	One of four kinds: is-VP, has-VP, active-VP, and passive-VP. An is-VP consists of is and either a predicate complement optionally followed by and and another predicate complement or the word either, a predicate complement, zero or more occurrences of a comma and a predicate complement, an optional comma, the word or, and a predicate complement. A has-VP consists of either has or does not have, a simple NP, the word as, and either a relational noun or a functional noun. An active-VP consists of either a singular verb or does not and an infinitive verb, either zero, one, or two simple NPs, and zero or more prepositional phrases. A passive-VP consists of is, an optional not, a past participle, an optional simple NP, and zero or more prepositional phrases.	`SimpleVP = "is", (PredComp, ["and" PredComp] | "either", PredComp, {",", PredComp},[","], "or", PredComp; | ("has" | "does not have"), SimpleNP, "as", (RelationalNoun | FunctionalNoun); | (VerbSing | "does not", VerbInf), [SimpleNP, [SimpleNP]], {PrepositionalP} | "is", ["not"], VerbPastPart, [SimpleNP], {PrepositionalP};`
Term	Zero or more adjectives followed by either a relational noun, an optional variable, the word of, and a list; or a noun, an optional variable, and an optional postmodifier.	`Term = {Adjective}, ( RelationalNoun, [Variable], "of", List | Noun, [Variable], [Postmodifier]);`
Text	A sequence of any number of headings and extended sentences. An empty text consisting of none is permitted.	`Text = {ExtendedSentence | Heading};`
Universal noun phrase	Either one of the words everything or everyone followed by an optional variable; or the word every followed by a term.	`UniversalNP = ("everything" | "everyone"), [Variable] | "every", Term;`
Verb phrase	One of four options: (1) a simple verb phrase, optionally followed by and and another simple verb phrase; (2) the word either followed by one or more simple verb phrases separated by commas followed by an optional comma, the word or, and a simple verb phrase; (3) the word is followed by a predicate complement, the word and, and a predicate complement; or (4) the words is either followed by a predicate complement, zero or more occurrences of a comma and a predicate complement, and finally an optional comma, the word or, and a predicate complement.	`VerbP = SimpleVP ["and", SimpleVP] | "either", SimpleVP, {",", SimpleVP}, [","], "or", SimpleVP | "is", PredComp, "and" PredComp | "is", "either", PredComp, {",", PredComp}, [","], "or", PredComp;`
WH-phrase	Either one of the words who, what, or when or the word which followed by a term.	`WhPhrase = ("who" | "what" | "when") | "which", Term;`

In general, CLCE offers some useful features; namely, its tight bidirectional linkage to FOL. It also permits user-defined expansion of the lexicon and claims to be able to perform error-checking because of the ability to do translations in both directions. On the down side, the expressiveness of CLCE is limited to FOL. It cannot express anything from natural language that would require higher-order logics, such as modal logic. The set of reserved and in-built predicates do not reflect a conceptualisation based on the requirements of a particular domain or an attempt at a set of universal semantic primitives and thus is, in some sense, an implementation language only.

2.4 Natural Semantic Metalanguage

2.4.1 Introduction

Natural Semantic Metalanguage (NSM) is a highly constrained metalanguage based on a relatively small set of approximately 60 English words and a prescribed syntax. The primary goal of NSM is to provide an irreducible set of *indefinibilia*, explicitly listed basic notions which constitute universal semantic primitives. This is motivated by the desire to define a well grounded metalanguage for the description of concepts encoded by human language.

In this survey we look at the theoretical stance of the approach, then present the semantic primes and undertake some preliminary analysis similar to that for the previous CNLs.

2.4.2 Theoretical stance

According to Wierbicka ([11] p. 1-34) an adequate explanatory account of natural language must take 'meaning' as a core object of investigation.

Wierbicka ([11] p. 9) claims two common widely-held assumptions with respect to language:

1. Definition of all words is possible
2. If a word is difficult to define, use a scientific-sounding word of Latin origin as a substitute.

Wierbicka ([11] p. 10) rejects both assumptions:

> The elements which can be used to define the meaning of words (or any other meaning) cannot be defined themselves; rather, they must be accepted as "indefinibilia", that is, as semantic primes, in terms of which all complex meanings can be coherently represented. A definition which attempts to explain the simple word *if* via the complex word *implication* flies in the face of the basic principle of sound semantic analysis put forward … by Aristotle ([16] p141a):
>
> First of all, see if he [the analyst] has failed to make the definition through terms that are prior and more intelligible. For the reason why the definition is rendered is to make known the term stated, and we make things known by taking not any random terms, but such as are prior and more intelligible … accordingly, it is clear that a man who does not define through terms of this kind has not defined at all.

For Wierbicka, the "absolute order of understanding" depends on semantic complexity ([11] p. 10):

> For example, one cannot understand the concepts of 'promise' or 'denounce' without first understanding the concepts of 'say', for 'promise' and 'denounce' are built upon 'say'. …
>
> When someone shows me a child who understands and can use the word *implication* but has not yet learned to understand and to use the word *if*, I will admit that everything is relative in semantics. Until such time, however, I will maintain that Aristotle was right, and that, despite all the interpersonal variation in the acquisition of meaning, there is also an "absolute order of understanding", based on inherent semantic relations among words.

One of the main assumptions of the semantic theory, and Wierbicka's semantic practice ([11] p. 11) is that 'meaning cannot be described without a set of semantic primitives; one can purport to describe meaning by translating unknowns into unknowns (as in Pascal's [17] p. 508 mock-definition "light is the luminary movement of luminous bodies"), but nothing is really achieved thereby.

Wierbicka ([11] p. 11) claims that '[w]ithout a set of primitives all descriptions of meaning are actually or potentially circular (as when, for example, *to demand* is defined as 'to request firmly', and *to request* as 'to demand gently'.

Wierbicka discusses the possibility of using arbitrary primitive terms, but rules this out for a semantic account of natural language on the basis that they 'will do little to advance our understanding of human communication and cognition' Wierbicka ([11] p. 11). On this point, she goes on to quote Leibniz:

> If nothing could be comprehended in itself nothing at all could ever by comprehended. Because what can only be comprehended via something else can be comprehended only to the extend to which that other thing can be comprehended, and so on; accordingly, we can say that we have understood something only when we have broken it down into parts which can be understood themselves. [18] p. 430 (translation Wierbicka).

For Wierbicka ([11] p. 11-12) 'Semantics can have an explanatory value only if it manages to "define" (or explicate) complex and obscure meanings in terms of simple and self-explanatory ones.'

Wierbicka suggests that seventeenth century philosophers, such as Descartes, Pascal, Arnauld, and Leibniz all believed that '[i]f a human being can understand any utterances at all … it is only because these utterances are built … out of simple elements which can be understood by themselves.' Wierbicka ([11] p. 11-12).

For Wierbicka, the empiricism basis of semantic analysis in modern (typological) linguistics provides a crucial basis for discovering 'fundamental concepts', moreover, Wierbicka claims ([11] p. 13) that '[i]n this sense, linguistics has a chance of succeeding where philosophical speculation has failed … and [she] does propose a complete (if hypothetical) table of fundamental human concepts capable of generating all other concepts'.

2.4.3 Proposal

Wierbicka ([11] p. 13ff) proposes a finite set of conceptual primitives universal to all languages and instantiated by words or bound morphemes.

Wierbicka ([11] p. 15) fully acknowledges this stance as the 'strongest universalist hypothesis', but tapers the point with two caveats:

1. the semantic systems embodies in different languages are unique and culture-specific
2. the presence of … lexicalised … universals does not mean perfect equivalence in language use.

However, Wierbicka ([11] p. 15) does claim that although '[w]hen we compare two, or more, languages we cannot expect to find identical networks of relationships. We can, none the less, expect to find corresponding sets of indefinables. … It is this (limited) isomorphism in the lexicon (and … grammar) that gives substance to the notion of universal semantic primitives.'

In addition to the set of semantic primitives, Wierbicka proposes a strict syntax for the formulation of semantic explications.

The finite set of semantic primitives amount to approximately 64 terms, as listed in Table 4.

Table 4: Proposed semantic primes (2002)

Substantives:	I, YOU, SOMEONE, PEOPLE, SOMETHING/THING, BODY
Determiners:	THIS, THE SAME, OTHER
Quantifiers:	ONE, TWO, SOME, ALL, MANY/MUCH
Evaluators:	GOOD, BAD
Descriptors:	BIG, SMALL
Intensifier:	VERY
Mental predicates:	THINK, KNOW, WANT, FEEL, SEE, HEAR
Speech:	SAY, WORDS, TRUE
Actions, events, movement, contact:	DO, HAPPEN, MOVE, TOUCH
Existence and possession:	THERE IS / EXIST, HAVE
Life and death:	LIVE, DIE
Time:	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT
Space:	WHERE/PLACE, HERE, ABOVE, BELOW; FAR, NEAR; SIDE, INSIDE; TOUCHING
"Logical" concepts:	NOT, MAYBE, CAN, BECAUSE, IF
Augmentor:	VERY, MORE
Taxonomy, partonomy:	KIND OF, PART OF
Similarity:	LIKE

The NSM website prescribes the use of these terms:

> A mere list is not sufficient, in itself, to identify the intended meanings, if only because many of these English words are polysemous (i.e. have several meanings), but only one sense of each is proposed as primitive. While it is claimed that the simplest sense of the exponent words can be matched across languages (i.e. that they are "lexical universals"), it is recognised that their secondary, polysemic meanings may differ widely from language to language.
>
> A fuller characterisation will indicate, for each proposed prime, a set of "canonical contexts" in which it can occur; that is, a set of sentences or sentence fragments exemplifying grammatical (combinatorial) contexts for each prime.
>
> When we say that a semantic prime ought to be a lexical universal, the term "lexical" is being used in a broad sense. A good exponent of a primitive meaning may be a phraseme or a bound morpheme, just so long as it expresses the requisite meaning. For example, in English the meaning A LONG TIME is expressed by a phraseme, though in many languages the same meaning is conveyed by single word. In many Australian languages the primitive BECAUSE is expressed by a suffix.
>
> Even when semantic primes take the form of single words, there is no need for them to be morphologically simple. For example, in English the words SOMEONE and INSIDE are morphologically complex, but their meanings are not composed from the meanings of the morphological "bits" in question. That is, in meaning SOMEONE does not equal "some + one" and INSIDE does not equal "in + side". In meaning terms, SOMEONE and INSIDE are indivisible.
>
> Semantic primes can also have variant forms (allolexes or allomorphs); for example, in English the word 'thing' functions as an allolex of SOMETHING when it is combined with a determiner or quantifier (i.e. this something = this thing, one something = one thing).
>
> Exponents of semantic primes may have different morphosyntactic characteristics, and hence belong to different "parts of speech", in different languages, without this necessarily disturbing their essential combinatorial properties.
>
> All these factors mean that testing the cross-linguistic viability of the proposed lexical primes is no straightforward matter. It requires rich and reliable data, and careful language-internal analysis of polysemy, allolexy, etc. Cross-linguistic testing of this kind is still in progress, and it is too early to be definitive about the outcome. But to date no convincing evidence has come to light which would disconfirm the universal status of any of the proposed semantic primitives. In general, therefore, the prospectus seems promising.

2.4.4 Reviews

The Natural Semantic Metalanguage (NSM) is an approach to semantic analysis based on reductive paraphrase (that is, breaking concepts/words down into combinations of simpler concepts/words …) using a small collection of semantic primes. The semantic primes … are believed to be atomic, primitive meanings present in all human languages. The concept has roots in the 17th century projects for ideal languages and the 18th century alphabet of human thought of René Descartes and Gottfried Leibniz.[7]

---

[7] http://en.wikipedia.org/w/index.php?title=Natural_semantic_metalanguage&oldid=-151177845

Riemer presents the Natural Semantic Metalanguage (NSM) framework of Wierzbicka and co-workers. Riemer sees NSM as a model of definitional enterprise, presented as a "refinement" on modern dictionaries. In proposing universally intertranslatable and indefinable semantic primitives, the NSM meant to evade charges of ethnocentrism, by developing a maximally culture-neutral methodology and objective terminology. Instead, it fell into semantic primitives that are not to be found in each culture ([19]).

A word in one language may be translated by a modified phrase in another language, for example, as promoted by meaning-based approaches to translation. The importance of this for semantic theory is that NSM theory claims that there is a set of universal, primitive concepts that can be expressed in any language. These concepts, the theory claims, being expressed in any language, can be used to formally define any concept in any language ([20]).

2.4.5 Analysis

From the conceptual framework identified in §1.3, NSM can be classified as attesting the following tendencies:

1. **analytic** vs. synthetic
2. **descriptive** vs. prescriptive
3. multiplicative vs. **reductionism**
4. formal vs. **informal**
5. **realism** vs. nominalism
6. **endurantism** vs. perdurantism
7. quantitative vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. **psychological** vs. non-psychological
10. **social** vs. non-social

Some important features of NSM for CNL are:

1. Proponents of NSM claim to be able represent any and all linguistically expressed concepts with NSM. This means that it should be able to function as an interlanguage for defining terms between humans either across languages or within the one language, and should be able to be mapped to concepts represented in a machine language.
2. Although extremely reductionist, that it takes a rigorous approach to the analysis of meaning, such that monosemy, polysemy, homophony, and other semantic relations will be carefully treated.
3. NSM claims that all very fine meaning distinctions in natural language can, and should, be described and represented by the formalism.
4. NSM does not avoid a treatment of relative terms of natural language such as quantifiers, e.g., 'more' and 'very'. These pose difficulties for formal languages and so careful and rigorous treatment by NSM can offer potential for 'formal' treatments.

We now compare ASD-STE100 and NSM. The following primes do not appear in ASD-STE100: I, SOMEONE, PEOPLE, THING, BODY, BIG, WANT, WORD(S), TRUE, DIE, MOMENT, MAYBE, LIKE.

We present in tabular form (Table 5) a comparison of keywords from the ASD-STE100 dictionary with corresponding terms in NSM. This gives a means of assessing the degree of overlap in use and intensional meaning between the primes and the ASD-STE100 lexicon.

Column one presents all the words which match the NSM prime. The keywords in capitals are those prescribed in ASD-STE100 with its prescribed part of speech and permissible forms. Words proscribed in ASD-STE100 but occurring in NSM are written in lowercase.

Column two presents the ASD-STE100 'Assigned Meaning' (AM). As indicated for Table 1, this is a loose indicator of the term's meaning. In the case of proscribed words, the cell indicates the preferred term, cf. 'want'. There are no definitions of the NSM terms as they are claimed to be primes and as such undefinable. (See, however, Wierzbicka [11] for extensive discussion of each prime.)

Column three presents the ASD-STE100 'Approved Example'. As for Table 1 these examples are partially successful in illustrating the intended meaning sometimes deficient from the AM. These first three columns repeat verbatim the information in ASD-STE100

Column four contains comments on problems concerning interpretation of meaning.

Table 5: Comparison of ASD-STE Keywords with NSM primes

Keyword (part of speech)	ASD-STE100 Assigned Meaning/ directed to USE	ASD-STE100 Approved Example	Comments re NSM
YOU (pn)	Related to the "user"	You can continue the test if the pressure is sufficient	NSM
SOMETHING (pn)	A "thing" that is not "determined" or specified	If something unusual occurs, do a bite test.	NSM
THIS (pn)	Function word that shows the person or "thing" referred to	Refer to "testing and fault isolation". This tells you the parts to replace.	NSM
SAME (adj)	Agrees in all details	The adjustment of the left and right wing flaps must be the same.	NSM
OTHER (adj)	Not the same as that given "before"	Connect one of the studs to the ground and the other studs to the pins.	NSM
OTHER (n)	That which is not the same as that given "before"	Connect one plug, then connect the other.	NSM
ONE (pn)	That refers to a "single" person or object. NOTE: Also a TN.	If the label is not there, attach a new one.	NSM
TWO (TN)		Do this procedure two times	NSM
SOME (pn)	Related to a quantity not specified	These bolts are shorter than the others	NSM
ALL (adj)	Full quantity	Drain all the fluids.	NSM
MANY (adj) (MORE, MOST)	Of large number NOTE: If possible, give the range, number, or quantity.	Many operators tell us that there is hydraulic leakage.	NSM
MUCH (adj) (MORE, MOST)	Of "relatively" large dimension, value, quantity, or volume	Do not apply too much pressure.	NSM
GOOD (adj) (BETTER, BEST)	This is satisfactory.	A XXXX is a good indication of this type of damage.	NSM

BAD (adj) (WORSE, WORST)	Not satisfactory. NOTE: Do not use this word if a more "specific" word is available.	For parking in bad weather conditions, refer to XXXX.	NSM
LARGE (adj) (LARGER, LARGEST)	More than average (TN) in dimension, quantity, or capacity.	Clean your skin with a large quantity of clean water.	NSM ► BIG
SMALL (adj) (SMALLER, SMALLEST)	Less than average (TN) in dimension, quantity, quality, or capacity	When you refuel the aircraft, a small quantity of fuel comes out of the vent line.	NSM
VERY (adv)	To a high "degree"	Add the oil very slowly.	NSM
THINK (v), THINKS, THOUGHT, THOUGHT	To have an "opinion"	If you think that the water will freeze, add antifreeze.	NSM
KNOW (v) KNOWS, KNEW, KNOWN	To be sure of data, to have data "ready" to use	If you know the clearance, you can calculate the washer thickness.	NSM
NECESSARY (adj)	That must be	If necessary, turn the fitting	NSM ► WANT
want (v)	NECESSARY (adj)	Put in the correct altitude, if it is necessary.	STE ► NECESSARY
FEEL (v) FEELS, FELT, FELT	To touch to find	You can feel the click in the handle when the torque wrench breaks.	NSM
SEE (v), SEES, SAW, SEEN	To know with the eyes (TN)	Make sure that you can see the cable through the inspection hole.	NSM, but !!! 'seeing is believing', cf. Turpin et. al. 'know with stomach'
HEAR (v), HEARS, HEARD, HEARD	To know by sound in the ear (TN)	Make sure that you hear the audio signals in all the crew headsets.	NSM, but !!! 'seeing is believing', cf. Turpin et. al. 'know with stomach'
SPEAK (v), SPEAKS, SPOKE	To use your voice to tell	Speak into the microphone.	NSM ► SAY
true (adj)	CORRECT, AGREE (v)	Make sure that the value shown on the indicator is correct.	

DO (v), DOES, DID, DONE	1. To complete a procedure, task or step	Do a functional test.	
DO (v), DOES, DID, DONE	2. Make	Do the repairs.	
DO (v), DOES, DID, DONE	3. As a helping verb a. as part of a negative "command" or "statement"	Do not move the XXXX if the unit does not operate.	
DO (v), DOES, DID, DONE	3. As a helping verb b. as part of a "question"	Does the light come on?	
happen (v)	OCCUR	If an accident occurs, get medical aid.	STE ► OCCUR
OCCUR (v), OCCURS, OCCURRED, OCCURRED	To be found, to "come to be," to "take place"	A sudden movement of the controls can occur.	NSM HAPPEN ► STE OCCUR
MOVE (v), MOVES, MOVED, MOVED	To change position or location	Make sure that the pointer moves freely.	NSM
TOUCH (v), TOUCHES, TOUCHED, TOUCHED	To be in "contact"	Make sure that the lock nut touches the spring.	NSM
THERE (pn)	Function word that starts a sentence	There are three jacking points on the fuselage.	NSM THERE IS, cf. STE THERE (adv)
IT (pn)	That "thing"	Carefully move the deflector unit down until it engages.	
exist (v)	BE	There are voltages which can kill in the main junction box.	STE ► THERE ARE cf. NSM THERE IS
HAVE (v), HAS, HAD, HAD	To "possess" as a part or quality	The system has an emergency warning device.	
LIVE (adj)	That includes explosive material NOTE: For electrical systems, USE: ENERGISE (v)	Some maintenance operations are not permitted on aircraft that have live ammunition	cf NSM LIVE
WHEN (con)	At the time that or during	When the slat movements stops, measure the travel.	NSM

TIME (n)	1. A "duration" that you can measure	The time between clicks must be 20 seconds.	NSM??
TIME (n)	2. An "occurrence"	Do step 2 three times.	NSM?? Which sense
now (adv)	AT THIS TIME	Do not tighten the nuts at this time.	NSM ►NOW
BEFORE (con)	That "precedes" a specified time, sequence, or operation	Bleed the system before you disconnect the components.	NSM??
BEFORE (pre)	That "precedes" a specified time, sequence, or operation	The bar moves down before 20 seconds.	NSM?? which sense
AFTER (con)	That follows a specified time, sequence, or operation	Does a functional test after you install the component.	NSM?? which sense
AFTER (pre)	That follows a specified time, sequence, or operation	The bar moves down after 20 seconds.	NSM?? which sense
PERIOD (n)	An unknown quantity of time	If you operate the engine with low fuel pressure for long periods, damage can occur.	NSM ► A LONG/SHORT TIME
SHORT (adj) (SHORTER, SHORTEST)	That has small length or "duration"	Attach the short arm of the bellcrank to the rod.	NSM ► A SHORT TIME
WHERE (con)	At, to, or in which location	Clean the area where you applied the sealant.	NSM
place (n)	POSITION, AREA	Make sure that the covers are in position	
POSITION (n)	The "attitude" or "setting" of something that you can adjust, or the "spot" or "site" where you put something.	Set the switch to the correct position.	
AREA (n)	A specified surface or location	Do not smoke in the work area.	NSM ► PLACE
HERE (adv)	In this position	The signal goes to the unit. Here the transformers change it into XXXX.	
ABOVE (pre)	In (or to) a position farther up then something NOTE: For other meanings, USE: MORE THAN	Lift the cylinder above its installed position.	NSM

BELOW (pre)	In (or to) a position farther down than something NOTE: For other meanings, USE: LESS THAN	The date is written below the cylinder neck.	NSM
FAR (adj) (FARTHER, FARTHEST)	At or to a "relatively" large distance	Remove the bolt that is farthest from the XXXX.	NSM
NEAR (adj) (NEARER, NEAREST)	At or to a "relatively" short distance	Put the tool on the near face of the installation.	NSM??
NEAR (pre)	"Approaching" in space or condition	Balance the elevator near its maximum limit.	
SIDE (n)	1. The specified surface or area of an object	Lubricate one side of the washer.	
SIDE (n)	2. A location or direction that has a relationship to a centre (TN) or a line (TN) of division (TN)	Push the unit out and then move it to the left side.	NSM
INNER (adj)	Nearer to the center (TN) of an object	The inner surface of the part is black.	NSM ► INSIDE
inside (n)	INNER (adj)	Clean the inner surface of the container.	
inside (n)	INNER (adj)	Paint the inner sides of the XXXX.	
inside (pre)	IN, INTO	Connect the supply union in the aircraft. Safety all components before you put them into the fuel tank.	
TOUCH (v), TOUCHES, TOUCHED, TOUCHED	To be in "contact"	Make sure that the lock nut touches the spring.	NSM ► TOUCHING
NOT (adv)	"Adverb of negation"	Do not let the pressure increase to more than 3000 psi	NSM
CAN (v), CAN, COULD	Helping verb that "means" to be "able" to	You can clean the drain holes with the clearing tool	NSM ???
POSSIBLY (adv)	That can occur	The length of the new rod is possibly incorrect.	NSM ► MAYBE

BECAUSE (con)	As a result of	Because the fluid level is incorrect, the system will not operate.	NSM
BECAUSE OF (pre)	As a result of	If you cannot remove a bolt because of corrosion, apply penetrating oil.	NSM
IF (con)	In the "event" that, "one the condition that", "in case of"	If you use a replacement fairing plate, cut the fairing plat to get the correct dimension.	NSM
VERY (adv)	To a high "degree"	Add the oil very slowly.	NSM
MORE (adj)	Refer to MANY/MUCH		NSM??
MORE (adv)	By a larger dimension, value, quantity, number of "degree"	If it is necessary to decrease the cable tension more, use the turnbuckle.	NSM??
kind (N)	TYPE	You can use two types of sealing compound.	NSM ► KIND OF
TYPE (n)	a SPECIFIED GROUP	Find the type and dimensions of the damage.	NSM ► KIND OF
PART (n)	1. A "constituent" of a machine (TN) or other equipment	Replace the damaged parts.	NSM ???
PART (n)	2. A "section" of a "whole"	Refer to part 2 for the applicable procedure.	NSM???
similar (adj)	Almost the same, equivalent	The two items have almost ehe same shape. Use material 11-001 or an equivalent material	NSM??
EQUIVALENT (adj)	Has the same properties, functions, or values	Materials XXXX and XXXX are equivalent.	NSM ► LIKE
EQUIVALENT (n)	Something that has the same properties, functions, or values	Use cleaning compound Ardrox 6025, or an equivalent.	

3. Descriptive Ontology for Linguistic and Cognitive Engineering

DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) has been developed as a reference module for a library of ontologies within the WonderWeb Project ([21], [22], [23]). DOLCE has been implemented using the OntoClean methodology. DOLCE makes the following choices.

1. **analytic** vs. synthetic
2. **descriptive** vs. prescriptive
3. **multiplicative** vs. reductionism
4. **formal** vs. informal
5. **realism** vs. nominalism
6. **endurantism** vs. perdurantism
7. **quantitative** vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. **psychological** vs. non-psychological
10. **social** vs. non-social

Although DOLCE has a natural language and human commonsense bias, it remains small and high-level. DOLCE is analytic, rather than synthetic, as it deals with top level concepts and relations, rather than with specific knowledge about the world.

DOLCE has a cognitive bias, as it aims at capturing the ontological categories underlying natural language and human commonsense ([22]) - it is therefore descriptive, not prescriptive.

DOLCE assumes that different entities can be co-located in space-time; it also allows entities that might be reducible to each other (e.g., points in space and regions of space) - it is therefore multiplicative, not reductionist.

Although DOLCE has a cognitive, natural language and human commonsense bias, and is aimed at capturing informal natural language expressions, it should be considered formal (despite the fact that DOLCE is neither prescriptive nor reductionist). DOLCE is formal, as there are KIF, FOL and OWL versions of DOLCE (moreover, the formal OntoClean framework of meta-properties has been employed as an ontological engineering tool).

DOLCE is an ontology of particulars, that does not classify universals (universals are excluded from the domain of quantification). Properties and relations are normally considered as universals, and even if they are not quantified over, they are used in the ontology to organise and characterise particulars. Nevertheless, DOLCE employs abstractions such as facts and sets, and therefore should be seen as a realist framework (although it is stated in [24] p.19 that "other examples of abstract entities (sets and facts) are only indicative").

DOLCE considers both endurants and perdurants, where endurants are wholly present at any time they are present. This precludes perdurantism proper, as perdurantism requires that

objects have distinct temporal parts. There are two types of parthood relations: a-temporal parthood (for perdurants) and time-indexed parthood (for endurants). Hence, DOLCE embraces endurantism.

DOLCE is both qualitative and quantitative, as it allows measures, but also includes qualitative concepts (mereotopological and temporal concepts). To some extent, DOLCE is functional, psychological and social, as it contains such concepts as process, accomplishment, mental object, social object, social agent and society.

DOLCE has been aligned with WordNet. The mapping between DOLCE and WordNet has been expressed in KIF.

The taxonomy of DOLCE has the following basic categories:

- Particular
    - Endurant
        - Physical Endurant
            - Amount of Matter
            - Feature
            - Physical Object
                - Agentive Physical Object
                - Non-agentive Physical Object
        - Non-physical Endurant
            - Non-physical Object
                - Mental Object
                - Social Object
                    - Agentive Social Object
                        - Social Agent
                        - Society
                    - Non-agentive Social Object
        - Arbitrary Sum
    - Perdurant
        - Event
            - Achievement
            - Accomplishment
        - Stative
            - State
            - Process
    - Quality
        - Temporal Quality
            - Temporal Location
        - Physical Quality
            - Spatial Location
        - Abstract Quality
    - Abstract
        - Fact
        - Set
        - Region
            - Temporal Region
            - Physical Region
            - Abstract Region

4. OpenCYC

OpenCyc is a publicly released component of Cyc, the commonsense knowledge base established by Doug Lenat from Cycorp ([25], [26], [27], [28]). Cyc is still under development. Cyc makes the following choices:

1. **analytic** vs. **synthetic**
2. **descriptive** vs. prescriptive
3. **multiplicative** vs. reductionism
4. **formal** vs. informal
5. **realism** vs. nominalism
6. **endurantism** vs. **perdurantism**
7. **quantitative** vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. **psychological** vs. non-psychological
10. **social** vs. non-social

It includes the entire Cyc ontology containing hundreds of thousands of terms, along with millions of assertions relating the terms to each other. It considers both analytic and synthetic facts.

The knowledge base is divided into microtheories each of which contains concepts and facts about particular areas of interest, with inheritance across some microtheories. Examples include “Movement”, “Intangible Things”, “Vehicles, Buildings and Weapons” and “Law”. The microtheories can provide alternate conceptualisations toward the same objects and so Cyc provides a multiplicative, rather than a reductionist, approach. For example, the Cyc web site ([26]) states,

> These assertions include both simple ground assertions and rules. Cyc is not a frame-based system: the Cyc team thinks of the KB instead as a sea of assertions, with each assertion being no more "about" one of the terms involved than another. … The microtheory mechanism allows Cyc to independently maintain assertions which are *prima facie* contradictory, and enhances the performance of the Cyc system by focusing the inferencing process.

Cyc expresses knowledge through the formal language, CycL, which is a LISP variant of a predicate calculus, not unlike KIF. Cyc constants include individuals, collections, truth functions (including connectives, predicates and quantifiers) and functions, which can be added to. It provides intensional expressivity a little beyond that first order logic and supports non-monotonic reasoning.

Cyc adopts a descriptive, rather than prescriptive, approach and incorporates a natural language subsystem that includes a lexicon, a syntactic parser and a semantic interpreter.

The ontology for Cyc is illustrated in Figure 4. At the uppermost level the concept "Thing" is defined as the "universal collection", and so a strong realist, rather than nominalist stance, is taken.

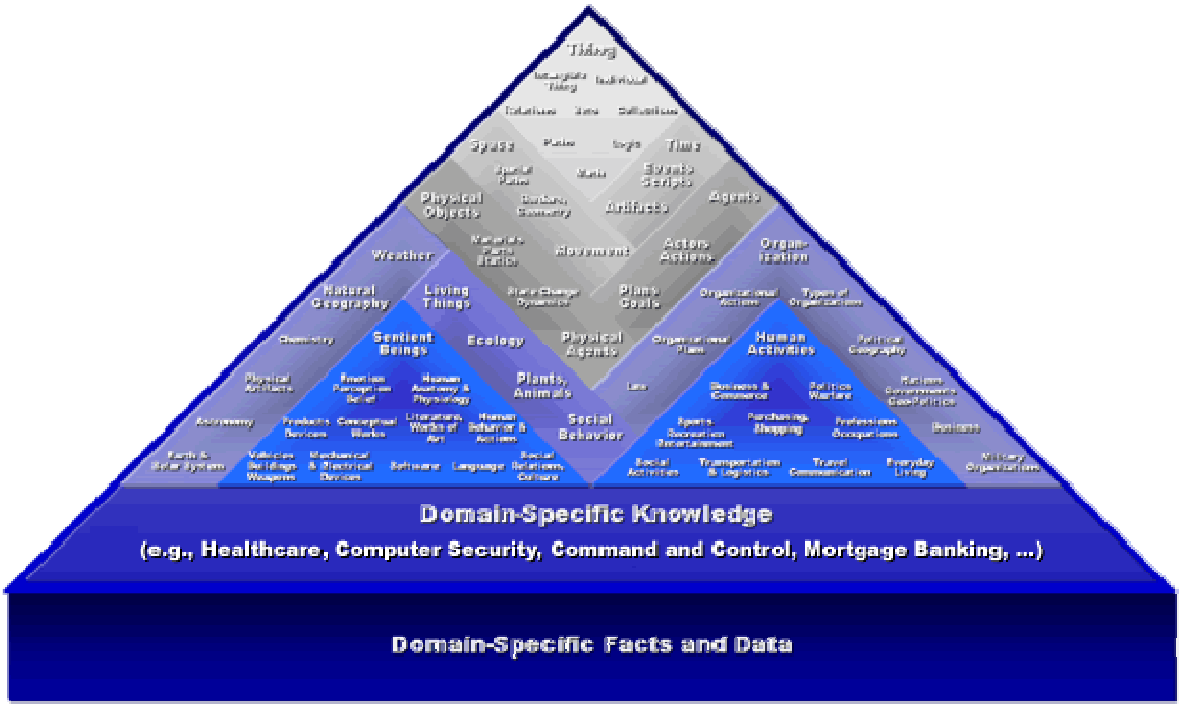

Figure 4: *Cyc Ontology*

The upper ontology for Cyc features in Figure 5 from [28].

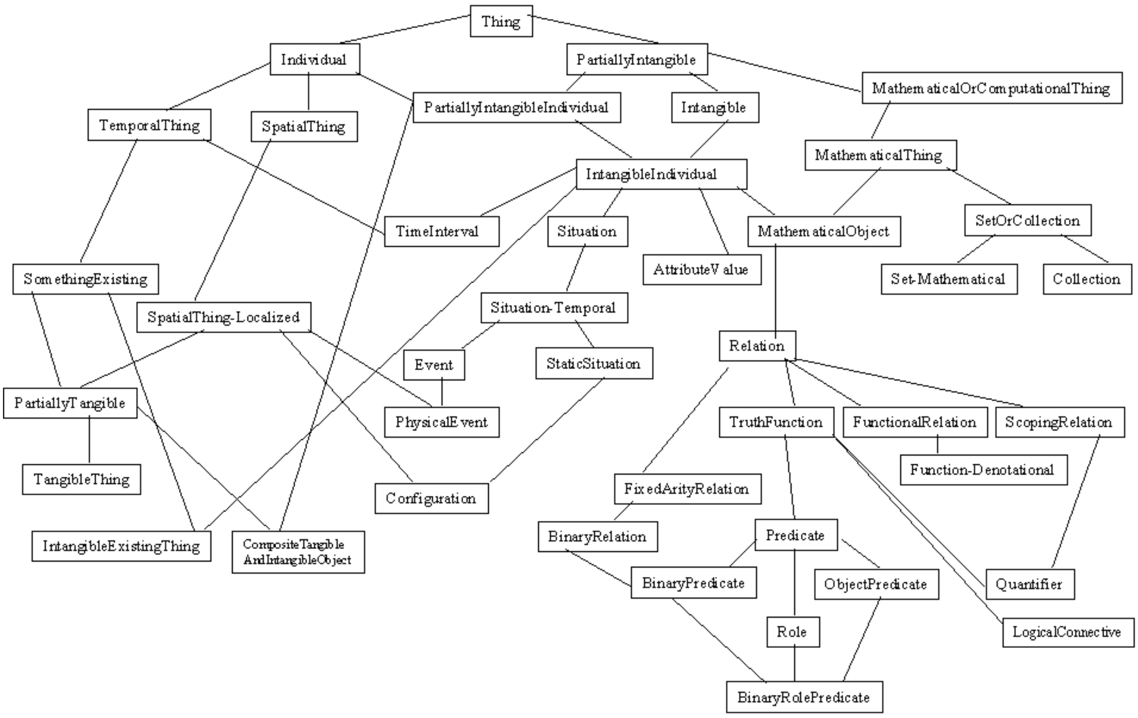

Figure 5: Cyc Upper Ontology

The Cyc upper ontology includes the concept "SomethingExisting" to provide endurantism and the concept "SituationTemporal" for perdurantism. From Figure 4 it is also apparent that the ontology clearly caters for social, psychological and some specific functional aspects. Qualitative and quantitative reasoning is supported. For example, temporal relationships can be represented qualitatively using concepts like "startsDuring" and quantitatively through numerically based date and time representations.

The lassez-faire nature of Cyc's descriptive and multiplicative approach makes it marry more readily with natural language approaches, but at times at the expense of logical constraints. The Cyc wikipedia article ([25]) offers the following criticisms of Cyc,

- The complexity of the system - arguably necessitated by its encyclopædic ambitions - and the consequent difficulty in adding to the system by hand
- Scalability problems from widespread reification, especially as constants
- Unsatisfactory treatment of the concept of substance and the related distinction between intrinsic and extrinsic properties
- The lack of any meaningful benchmark or comparison for the efficiency of Cyc's inference engine
- The current incompleteness of the system in both breadth and depth and the related difficulty in measuring its completeness

- Limited documentation
- The lack of up-to-date on-line training material makes it difficult for new people to learn the systems
- Contrary to claims there is no "open source" ontology available nor an "open source" version of assertions
- A large number of gaps in not only the ontology of ordinary objects but an almost complete lack of relevant assertions describing such objects
- An amateurish feel to the entire project that point to the pitfalls of large "open" projects or perhaps the commitment of its founder.

5. Mephisto

The Mephisto framework was first proposed in [2] and has been developed by Dale Lambert and Chris Nowak ([29]). It is designed to represent aspects of the military and national security domains. It makes the following selections.

1. **analytic** vs. synthetic
2. descriptive vs. **prescriptive**
3. multiplicative vs. **reductionism**
4. **formal** vs. informal
5. realism vs. **nominalism**
6. endurantism vs. **perdurantism**
7. **quantitative** vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. **psychological** vs. non-psychological
10. **social** vs. non-social

The Mephisto framework is based on the five layer model from [2] illustrated in Figure 6.

Social:	group, ally, enemy, neutral, own, possess, invite, offer, accept, authorise, allow.
Intentional:	individual, routine, learnt, achieve, perform, succeed, fail, intend, desire, belief, expect, anticipate, sense, inform, effect, approve, disapprove, prefer.
Functional:	sense, move, attack, attach, inform, operational, disrupt, neutralise, destroy.
Physical:	land, sea, air, outer_space, incline, decline, number, temperature, weight, energy.
Metaphysical:	exist, fragment, identity, time, before, space, connect, distance, area, volume, angle.

Figure 6: Mephisto Framework

The physical, functional and intentional layers were motivated by Dennett's ([30]) physical, design and intentional stances respectively in which he argues that individuals will seek to predict and explain an entity on the basis of naïve physics where possible, then on the basis of the entity's design if the physical stance is unsuccessful, and then on the basis of a cognitive stance toward the entity if the design stance is unsuccessful. To this Lambert ([2]) added a metaphysical layer below, a social layer above, and contemplated the nature of the relations that occupy each layer, with each layer reliant on relations from the layers below. The Physical Layer was subsequently renamed as the Environmental Layer. The Intentional Layer was subsequently renamed as the Cognitive Layer ([31]). Based *loosely* on the names of its layers Metaphysical, Physical, Functional, Intentional, Social, Nowak ([29]) introduced the term "Mephisto" for the conceptual framework.

The role of each layer is outlined in [31],

> The *Metaphysical Layer* introduces foundational concepts like existence, identity, space and time. This allows the machine to identify fragments of the environment of interest and to do so with respect to their spatiotemporal parts. The *Environmental Layer* introduces environmental properties and relations to the metaphysical parts. This allows the machine to ascribe attributes like temperature and weight to individuated parts, while identifying some parts as ocean, others as land, and so on. The *Functional Layer* considers the functionality of identified physical parts. Principal functional relations in a military context are the ability to sense, move, strike, attach (includes carry), inform and transform. These are sufficient to characterise: surveillance and reconnaissance, weapons, logistics, communications and engineering capabilities. The *Cognitive Layer* adds cognitive relations to the identified physical and functional spatiotemporal parts. The attribution of beliefs, intentions and other mental states is performed at the Cognitive Layer. Finally, the *Social Layer* introduces social constructs between the cognitive individuals. Concepts like authority and enemy prevail at the Social Layer.

Consequently Mephisto considers functional, psychological and social aspects.

As an ontological framework, Mephisto is prescriptive, rather than descriptive, in that it attempts to provide a philosophically well grounded approach, rather than a natural language oriented conceptualisation. Mephisto is also a reductionist, rather than a multiplicative approach, in that it seeks to identify a small number of primitive terms that are sufficient to account for the five layers in the intended context. The Mephisto focus is on meaning, and hence is on an analytic rather than synthetic knowledge. As outlined in [32], the primitive concepts are to be formalised in a logic to make the meanings of those terms precise and then implemented within a machine with a logical reasoner. This will allow a machine to reason meaningfully with those concepts. The framework marries both qualitative and quantitative information. Mephisto takes a strong stance on nominalism and perdurantism, by conceptualising all objects as processes composed of "bare matter".

Mephisto is presently only partially formalised, so one drawback is that additional development is required. Its strength is its philosophical rigour and its emphasis toward developing sound and complete formal theories where possible. This makes it amenable to computational implementation and provides clarity of interpretation, but at a price of being further removed from natural language, which is more suited to a descriptive and multiplicative approach, rather than Mephisto's prescriptive and reductionist style.

6. Suggested Upper Merged Ontology

The Suggested Upper Merged Ontology (SUMO) is an upper ontology which was developed by Ian Niles and Adam Pease in 2001 and is now maintained by the Teknowledge Corporation. SUMO was developed to facilitate data interoperability, information search and retrieval, automated inference and natural language processing ([33]). It is also one of three candidates under consideration for the establishment of the IEEE Standard Upper Ontology (SUO) ([34] p4-1). It is now owned by the IEEE and is freely available under the GNU General Public Licence ([35]).

As its name suggests, the development of SUMO was based on the merging of different extant ontologies with freely available content including ([33], [34] p4-1 and [35]):

- John Sowa's upper ontology;
- Russell and Norvig's ontology;
- Process Specification Language (PSL);
- Casati and Varzi's theory of holes;
- Allen's temporal axioms;
- The relatively non-controversial elements of Smith and Guarino's respective mereotopologies;
- The KIF formalisation of the Core Plan Representation (CPR);
- The ontologies available on the Ontolingua server maintained by Stanford University's Knowledge Systems Laboratory;
- The ontologies developed by the ITBM-CNR;
- Some of the spatial relations in an unpublished paper by Iris Tommelein and Anil Gupta entitled "Conceptual Structures for Spatial Reasoning"; and
- A "Structural Ontology" proposed by David Whitten and substantially revised and extended by Chris Menzel.

During the early development phase of SUMO, Sowa's upper ontology and Russell and Norvig's ontology, which were both very compact and contained a significant amount of overlapping content, were first merged into a single upper ontology. This was then used as the foundation for aligning all of the other ontologies ([33]) into a single consistent ontology. It is now quite a mature ontology and has been applied in the disciplines of linguistics, knowledge representation and reasoning ([36] p6 and [37]). SUMO is the largest free formal ontology available; its current version 1.75 features 20,000 terms and 60,000 axioms ([35]). It is written in a version of KIF known as SUO-KIF ([33]). A system called KSMSA, which supports visual editing and displaying of the ontologies, is also available. Other resources and information about SUMO are available from the Ontology Portal website [35].

A subset of the top level categories in SUMO is portrayed in Figure 7 and a stylised version of its structure is illustrated in Figure 8 ([35]). Figure 8 also highlights the existence of the mid-level ontology (MILO) and a variety of domain ontologies that have been developed as adjuncts to SUMO. The most current list of domain ontologies includes ([35]):

- Communications;
- Countries and regions;
- Distributed computing;
- Economy;
- Finance;
- Engineering components;
- Geography;
- Government;
- Military (general, devices, processes, people);
- North American Industrial Classification System;
- People;
- Physical elements;
- Transnational issues;
- Transportation;
- Viruses;
- World airports (A-K and L-Z);
- WMD; and
- Terrorism (available upon request).

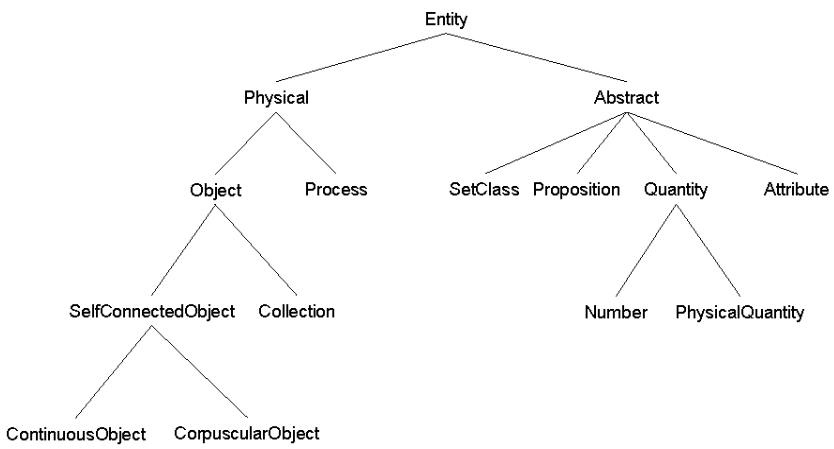

Figure 7: A Subset of Top Level Categories in SUMO [34]

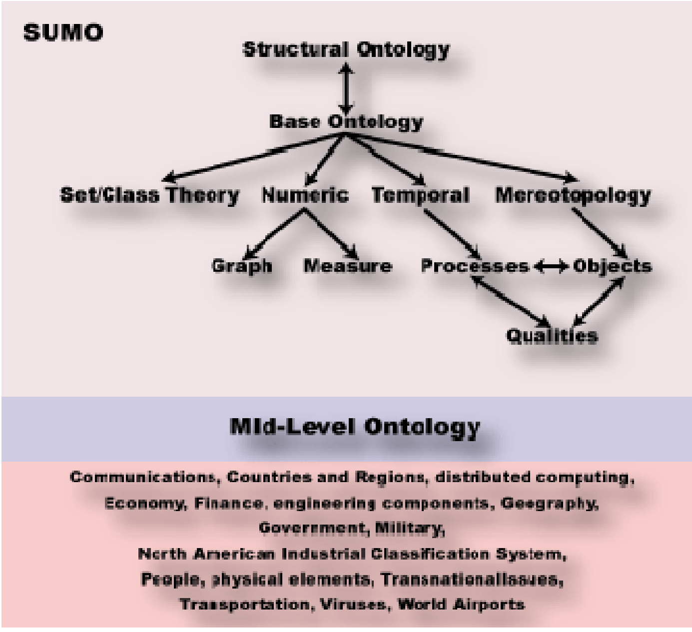

Figure 8: A Stylised View of the Suggested Upper Merged Ontology Structure along with its Mid-Level Ontology (MILO) and a Selection of its Domain Ontologies [35]

Other features of SUMO that add to its utility include:

- Mappings from all of the integrated SUMO ontologies into WordNet 3.0 ([38],[39] and [40]);
- Existence of theorem provers for reasoning with SUMO eg Vampire and the FOL reasoning scheme of Baumgartner *et al* ([41] and [42]); and
- The development of a new domain ontology for SUMO for Information Technology Security Evaluation ([43]).

An ontology called OntoSensor for symbolically expressing information extracted from sensor data is also in its early stages of development. While it is based on definitions, concepts and properties in part from SUMO, OntoSensor is not a component ontology of SUMO. However, it is conceivable that it could be integrated into SUMO at some later date ([44]).

As an upper ontology, SUMO may be classified as follows:

1. **analytic** vs. synthetic
2. descriptive vs. **prescriptive**
3. **multiplicative** vs. reductionism
4. **formal** vs. informal
5. **realism** vs. nominalism
6. **endurantism** vs. **perdurantism**
7. **quantitative** vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. **psychological** vs. non-psychological
10. **social** vs. non-social

Amplifying Comments:

- **Analytic** vs. synthetic – SUMO is aimed at developing an analytic upper ontology.
- Descriptive vs. **prescriptive** – Given the size of SUMO, it is difficult to state categorically that SUMO is wholly prescriptive, although with the huge number of concepts that it possesses, this must be the case for most of the ontology. For example, it possesses conceptualisations for set theory and relations which are prescriptive in nature ([34] p4-2).
- **Multiplicative** vs. reductionism – It is stated in [34] p4-2 that SUMO is multiplicative.
- **Realism** vs. nominalism – It is stated in [34] p4-1 that SUMO contains elements of realism.
- **Endurantism** vs. **perdurantism** – SUMO possesses conceptualisations for both objects and processes ([34] p4-2), so is capable of representing both endurant and perdurant aspects of entities. Some of the processes that SUMO is capable of capturing include dual object processes, internal change, shape change, intentional processes, guiding, social interaction, motion and transportation ([45]).
- **Quantitative** vs. **qualitative** – SUMO possesses both quantitative and qualitative means of describing at least some entities. For example, space and time can be expressed both quantitatively and qualitatively in SUMO.
- **Functionalism** vs. non-functionalism – Some of the concepts for handling objects and processes should together provide a means of establishing some functional concepts ([45]).
- **Psychological** vs. non-psychological – SUMO possesses a concept of psychological processes ([46] pp 215, 218, 224, 246-250, 268, 281, 386-388).
- **Social** vs. non-social – SUMO possesses several social concepts such as social roles, social interactions and social unit ([46] pp 41, 239, 269-270, 276-277, 333, 341, 343, 365-366).

7. Basic Formal Ontology

The BFO (Basic Formal Ontology) ([49], [50], [51], [52]) contains two ontologies, SNAP (endurants) and SPAN (perdurants), with some links between them (e.g., participation relation links endurants to perdurants). Additionally, BFO allows different granularity views on reality. The following choices are made.

1. **analytic** vs. synthetic
2. descriptive vs. **prescriptive**
3. multiplicative vs. **reductionism**
4. **formal** vs. informal
5. **realism** vs. nominalism
6. **endurantism** vs. perdurantism
7. **quantitative** vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. **psychological** vs. non-psychological
10. **social** vs. non-social

BFO is analytic, rather than synthetic, as it deals with top level concepts and relations.

BFO (and DOLCE) is, according to [34], descriptive and multiplicative. However, according to [50] DOLCE is multiplicative, but BFO is not: "the vase and the clay are not to be genuinely distinguished in BFO". Furthermore, according to [53]:

> BFO commits to a reductionist stance w.r.t. co-localised entities. BFO assumes that reality and its constituents exist independently of our (linguistic, conceptual, theoretical, cultural) representations thereof.

BFO has been implemented in OWL and therefore is formal. BFO also employs some formal theories, such as mereology and a theory of spatial location ([51]).

BFO admits both particulars and universals in the domain of quantification. It assumes reality of universals and is therefore a realism framework. In [51], p.11, Grenon says:

> A definite account of categorial membership ought to rest on a form of Aristotelian realism about universals. Universals account for categorial patterns in the world. They exist in their instances, and not without their instances.

BFO has a SNAP ontology of endurants and a SPAN ontology of endurants. As it allows both endurants and perdurants, it takes the endurantism stance.

BFO is both qualitative and quantitative. It is also to some extent functional, psychological and social ([40]).

A taxonomy of BFO basic categories appears below.

- Entity
  - Enduring Entity
    - Spatial Region
      - Free Portion of Space
      - Physically Bound Portion of Space
    - Dependent Entity (Tropes)
      - Quality, State, Power
      - Quasi-Quality, Quasi-State, Quasi-Power
    - Independent (Substantial) Entity
      - Substance
      - Aggregate of Substances
      - Fiat Part of Substance
      - Boundary of Substance
      - Quasi-Substance
  - Concrete Entity
    - Spatio-Temporal Region
    - Processual Entity
      - Process
      - Aggregate of Processes
      - Fiat Part of Process
      - Instantaneous Temporal Boundary of Process
      - Quasi-Process

Figure 9: A Taxonomy of BFO Basic Categories

8. Object-Centred High-Level Reference Ontology

The Object-Centered High-Level Reference Ontology (OCHRE) is an upper ontology developed by Luc Schneider from the Department of Philosophy at the University of Geneva in 2003. While it was developed independently of other upper ontologies, it has since been incorporated with DOLCE and BFO as part of a single library under the WonderWeb project (see [24]).

Conceptually, it covers top-level notions related to object, attribute, event, parthood, dependence and spatio-temporal connection ([54] p1). More precisely, it incorporates formal theories of mereology (parthood), foundations, similarity, topology, properties (attributes), relational properties and eventualities ([24] pp 165-177). The FOL axiomatisation of OCHRE is outlined (and compared to DOLCE) in [54] and a KIF version of it appears in [24] pp 165-177.

Underpinning OCHRE is the taxonomy in Figure 10 for describing entities both in terms of their enduring aspects and their changing nature over time ([53] p.102).

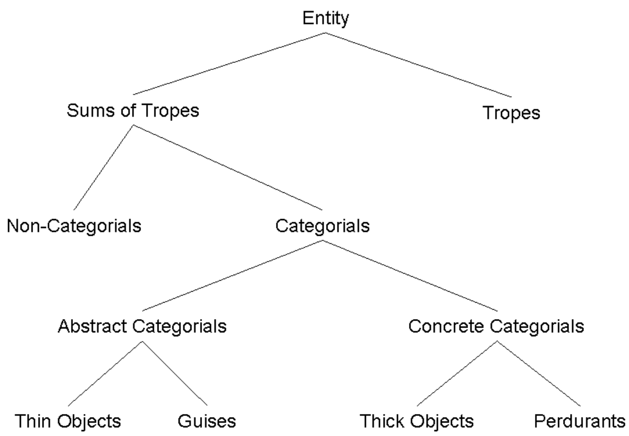

Figure 10: The OCHRE Taxonomy [53], p. 102

OCHRE is an ontology of "particulars", that is concrete individuals (entities) that are qualitatively accounted for as bundles of simple (atomic) individual features i.e. non-repeatable properties and relations. The mereological atoms which describe the individual characteristics, upon which entities and their spatio-temporal facets are defined, are referred to as Tropes. Tropes may be bundled (or summed) together to build up representations of entities. These Sums of Tropes may be further specialised as Non-Categorials, which comprise arbitrary aggregations of Tropes, and Categorials. Categorials may be even further distinguished as Abstract or Concrete on the basis of their spatio-temporal extension. Concrete Categorials comprise the Thick Objects, which are aggregations of Tropes that are extended in time and space, and Perdurants, which are successions of Thick Objects. Abstract Categorials on the other hand comprise the Thin Objects, which may be thought of as a core of essential enduring (ie spatio-temporally invariant) properties of an entity, and Guises, which are sums of Thin Objects and the Tropes dependent on them. In this way, the representation of entities may be regarded as having both elements of Endurantism and Perdurantism. In OCHRE change is conceptualised as a succession of Thick Objects which are linked by a common Thin Object ([53] pp 201-202 and [24] pp 44-45)

The following example, drawn from [53] p. 101, serves to illustrate these ideas.

> The Tropes of a ripening tomato are its color, its mass, its shape, etc. The change of a ripening tomato just pertains to different Thick Objects representing the tomato and its Tropes. That means the Thick Objects are wholes centered around the bundle of core characteristics, e.g. the tomato's DNA, represented by a Thin Object. That one speaks of the same object through change is grounded in the existence of Thin Objects.

As an upper ontology, OCHRE may be classified as follows:

1. **analytic** vs. synthetic
2. descriptive vs. **prescriptive**
3. multiplicative vs. **reductionism**
4. **formal** vs. informal

5. realism vs. **nominalism**
6. **endurantism** vs. **perdurantism**
7. quantitative vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. psychological vs. **non-psychological**
10. social vs. **non-social**

Amplifying Comments:

- Descriptive vs. **prescriptive** – Conceptualisations in OCHRE have some basis in commonsense, but the view taken during OCHRE's development was that any attempt to read off an ontological commitment from observable human endeavour, either linguistic or otherwise, would be problematic because commonsense is not constant ([24], pp 42-43). As such, OCHRE "aims at combining descriptive adequacy for commonsense with formal economy in the basic categories and their axiomatisation [*ibid*. p.43]". However, it is certainly more prescriptive than descriptive because the distinction between Thin Objects and Thick Objects does not accord with human commonsense ([53], p102).
- Multiplicative vs. **reductionist** – While the philosophy behind OCHRE is to reject the multiplicative approach ([24], p 43), some observers have reported that it is unclear if it truly adopts a reductionist stance ([53], p102).
- Realism vs. **nominalism** – Some aspects of realism may be captured within OCHRE, such as colour for example, through the use of Tropes. However, it is heavily slanted towards a nominalist view of the world through its focus on objects, mereology and topology. Furthermore, OCHRE's "reductionist" stance does not support a "possible worlds" view. Nor does it support other realist abstractions such as propositions and sets ([24], p 43).
- **Endurantism vs. perdurantism** – As already outlined, the combined use of Thin and Thick Objects to describe entities is somewhat of a hybrid of the endurant and perdurant philosophies.
- Quantitative vs. **qualitative** – In both [24] p.50 and [53] p.101, it is stated that OCHRE provides qualitative accounts of both objects and events.
- **Functionalism** vs. non-functionalism – To some extent, OCHRE supports functionalism via the concept of "participation" which is defined as a special case of parthood. A Thin Object x participates in a process y if and only if x is the substrate of an event that is part of y ([54] p12).
- Psychological vs. **non-psychological** – In its current form, OCHRE does not possess a psychological conceptualisation.
- Social vs. **non-social** – In its current form, OCHRE does not possess a social conceptualisation.

9. MIP Data Models

The Multilateral Interoperability Programme (MIP) [55] is concerned with overcoming the problems of interoperability between distinct National Command and Control Systems (C2S). Part of the MIP solution is the development of a common data model for representing C2 Information intended for exchange between nations. As such the data model represents a loose form of ontology, justifying its consideration herein.

9.1 C2IEDM

MIP Baseline 2 was released in September 2006, and contained the **Command and Control Information Exchange Data Model** [56] as a core element. Currently at version 6.15e, the C2IEDM is central to the capability that MIP provides, which is the exchange of C2 information. C2IEDM is based originally on the ACCTIS-published LC2IEDM (Land C2 Information Exchange Data Model), also known as Generic Hub 5 (GH-5).

The C2IEDM is a common data model shared between participants in a MIP solution. All information from a C2IS that is to be exchanged to a partner system through the MIP solution needs to be mapped into a C2IEDM repository. Exchange mechanisms share this information from the repository to another MIP-enabled system. The C2IEDM specification describes means of storing information about objects such as organisations, people, material, facilities and equipment, as well as geographic features and geospatial information (e.g. forward lines of battle, *et cetera*). The data model is capable of storing (and sharing) meteorological information.

Geospatial data and location information for features and objects may be stored in absolute terms or relative to some origin point. All coordinate information is in the World Geodetic System 1984. Reporting data and the organisations responsible for the origin of the report is recorded and associated to the information that has been reported.

C2IEDM is extensible, accommodating the possibility of a specific C2IS that encapsulates C2 concepts that have not been captured through the original MIP Requirements gathering process. Such extensions can make C2IEDM better fit an individual nation's needs, however, these extensions are not sharable outside of the specific nation that implements them. The MIP specification and business rules would cause novel shared information types to be rejected by a standard MIP solution system.

9.2 JC3IEDM

The MIP Block 3 effort is currently working to produce Baseline 3[8]. The data model for this Baseline is called **Joint C3 Information Exchange Data Model** (JC3IEDM). Additional capability is being introduced into the model such that there are concerns regarding the backward-compatibility of the model with the C2IEDM of Baseline 2. The MIP community is looking to address this in future Blocks by putting in place rules governing the evolution of

[8] JC3IEDM v3.1a was published to the MIP Members pages 12/3/2007, but is dated 16/2/2007.

the data model to provide as much backwards compatibility as possible in place between Block 3 and further evolutions of the solution.

Block 3 also introduces the possibility of an additional exchange mechanism, based on the eXtensible Markup Language (XML). The XML Exchange Mechanism (XEM) is as yet unformalised, but groundwork is being laid in Annex O of the current JC3IEDM specification which deals with Extensible Markup Language (XML) Reference Schemas and Implementation Guidance.

The JC3IDM comprises the following Independent Entities and defines their roles within the main body of the specification: ACTION, ADDRESS, AFFILIATION, CANDIDATE-TARGET-LIST, CAPABILITY, COMPONENT-HEADER-CONTENT, COMPONENT-TEXT-CONTENT, CONTEXT, RELATIVE-COORDINATE-SYSTEM, GROUP-CHARACTERISTIC, LOCATION, OBJECT-ITEM, OBJECT-TYPE, PLAN-ORDER, REFERENCE, REPORTING-DATA, RULE-OF-ENGAGEMENT, SECURITY-CLASSIFICATION, and VERTICAL-DISTANCE.

Individually named objects in the world are part of the OBJECT-ITEM classification, whereas OBJECT-TYPE refers to class objects. Implicit in the distinction is that the OBJECT-TYPE tends to be a static representation of the characteristics of an object such as the calibre of the main gun of a type of tank. This is contrasted by the OBJECT-ITEM, which is open to more dynamic changes, such as the call sign of the tank, or its operational status.

Figure 11 lays out the relationships between the independent entities, where a dot indicates "many" (specifically zero or more) and solid lines indicate many-to-many, whereas dashed lines indicate non-identifying one-to-many.

The OBJECT-ITEM and OBJECT-TYPE concepts are core to the model, and are broken down into a first-level sub classification as shown in the above figure. In Figure 12a, the circle with a single line is used to represent a complete subtyping relationship, permitting no further siblings of the explicated subtypes. A double line beneath the circle indicates a subtyping relationship which does not explicate all possible subtypes, as in the subtypes of action shown in Figure 12b. A full explanation of the IDEF1X notation is found in C3IEDM Annex K ([57]).

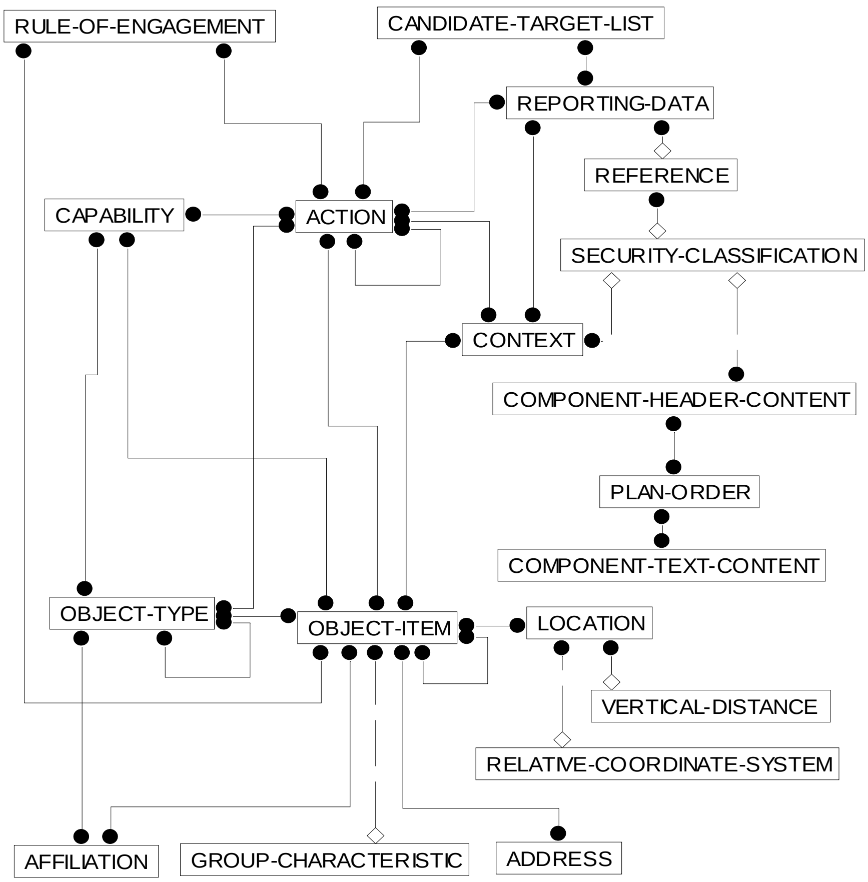

Figure 11: Independent Entities for Creating the Data Specification

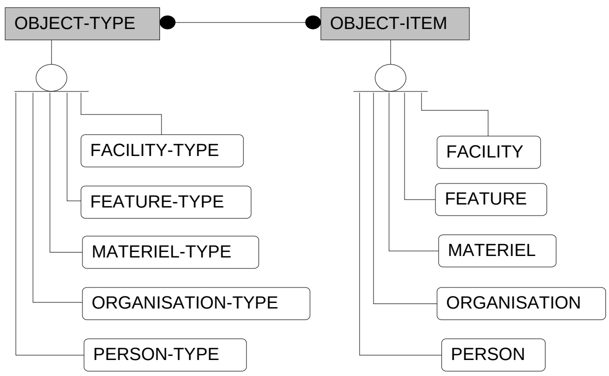

Figure 12a:First Level Subtyping of OBJECT-TYPE *and* OBJECT-ITEM

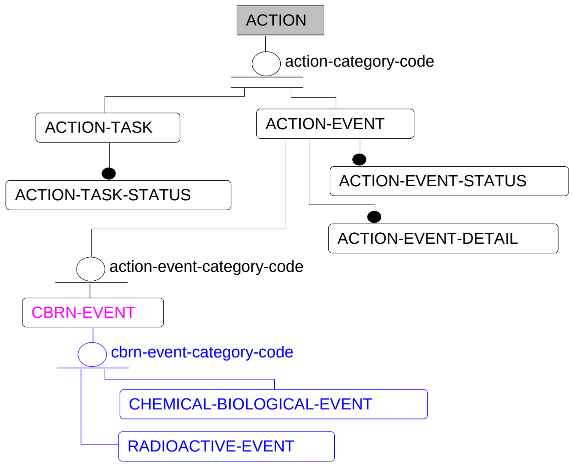

Figure 12b:(partial) ACTION *Subtype Structure*

JC3IEDM makes the following distinctions:

1. **analytic** vs. synthetic
2. **descriptive** vs. prescriptive
3. **multiplicative** vs. reductionism
4. **formal** vs. informal
5. realism vs. **nominalism**
6. **endurantism** vs. perdurantism
7. **quantitative** vs. qualitative
8. **functionalism** vs. non-functionalism
9. psychological vs. **non-psychological**
10. **social** vs. non-social

The analytic construction of the JC3IEDM is done through the collection of requirements for information exchange primarily from the Army community. There is no process of identification of the fundamental elements of the model and building up the rest of the concepts encompassed by the model from these basic building blocks. Instead, capturing the things in the domain of the C2 environment that particular nations desire to exchange between themselves is done on an almost *ad hoc* basis that has been elaborated through the evolution of the data model.

Although the MIP data models do not present themselves in a particularly natural language manner, being either logical or physical database tables, or an XML representation thereof (available only in JC3IEDM, not in its predecessor), the approach taken is more descriptive than prescriptive. There is no philosophical or scientific framework that underpins the concepts captured by the data model. As such the data models are also multiplicative, containing several relatively unrelated schemata of classification of the things in the world.

The model put forward in the MIP specifications is a formal one: it is encoded in an entity-relationship notation (IDEF1X) within the specifications, and is also published in ERWin Data Modeller format. The formal model is used as input to the XML tools suite to generate XML

Schema Document. It is possible that a more logically-oriented formal encoding could be derived from the base model, however, the paucity of captured semantics in the original model may make this a sparse hierarchical arrangement of concepts with little distinction between them.

The MIP data model can be characterised as nominalist, lacking any instantiation of universals, nor admission of the use of sets. Properties of objects may be used to record aspects of the things in the world and to some degree this is also possible for the classes of things (OBJECT-TYPE may also have properties), however, this is a highly proscriptive data model, permitting extension only in fulfilling the needs of a National C2S.

Entities encoded in the data models are endurant; even at the point where their state is changed from, say, *operational* to *destroyed* the entity remains in the database. Such changes in state do not alter the identity of the object. The MIP data models are highly quantitative; e.g. physical features and battlespace geometries are encoded as precise geographic entities and no mechanism for capturing the binary geospatial relationships between such entities is provided.

The MIP models are distinctly functional, in that equipment and organisations such as military units have a capability to perform certain functions. They are also able to “move” through reporting new locations for such things over time. The entities in the data models have no psychological attitudes explicitly encoded, however a very loose and informal means of capturing such could be done through intelligence reports and observations, as well as possibly through the commander’s intent attached to plans and orders represented in the system. These informal captures are available only in free text character information, rather than through specific and explicit entities in the model. From a social point of view, there is a limited representation available in the form of hostility codes of various military units, as well as hierarchies of command within military organisations.

10. Conclusions

In Figure 2 the Conceptual work package interfaces with the Language work package and the Formal work package, which in turn interfaces with the Computational work package.

From the natural language standpoint, the better fit is for the following to hold for a conceptualisation.

1. **analytic** vs. **synthetic**
2. **descriptive** vs. prescriptive
3. **multiplicative** vs. reductionism
4. formal vs. **informal**
5. **realism** vs. nominalism
6. **endurantism** vs. perdurantism
7. **quantitative** vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. **psychological** vs. non-psychological
10. **social** vs. non-social

From the formal/computational standpoint, the better fit is for the following to hold for a conceptualisation.

1. **analytic** vs. **synthetic**
2. descriptive vs. **prescriptive**
3. multiplicative vs. **reductionism**
4. **formal** vs. informal
5. realism vs. **nominalism**
6. **endurantism** vs. **perdurantism**
7. **quantitative** vs. **qualitative**
8. **functionalism** vs. non-functionalism
9. **psychological** vs. non-psychological
10. **social** vs. non-social

Thus there are tensions in the requirements, suggesting that no one solution will suffice. Cyc offers the best synthetic solution, but its analytic framework is less thought through. It would prove valuable if we need basic knowledge involving a large number of commonsense concepts.

Cyc's natural language extension and STE provide the two most descriptive formulations, and Cyc claims to provide a lexicon, a syntactic parser and a semantic interpreter. These represent the high end language options, together with Mel'cuk's Meaning Text (MT), which is not reviewed here. The progression to a computational solution with the Cyc natural language system would be via Cyc. The danger with this is that the multiplicative approach appears to have scope for inconsistent responses, depending upon which microtheories are invoked. The

progression to a computational solution with the STE or MT options, or a natural language framework based on the scenario directly, could be via NSM, which provides a descriptive and reductionist account. CLCE or PENG could be applied below this to translate to a logical format, or this could be considered directly. The formal options are DOLCE, SUMO, CYC and Mephisto. The first three are mature but multiplicative. The latter is immature but reductionist and more domain focused. Figure 13 characterises the options.

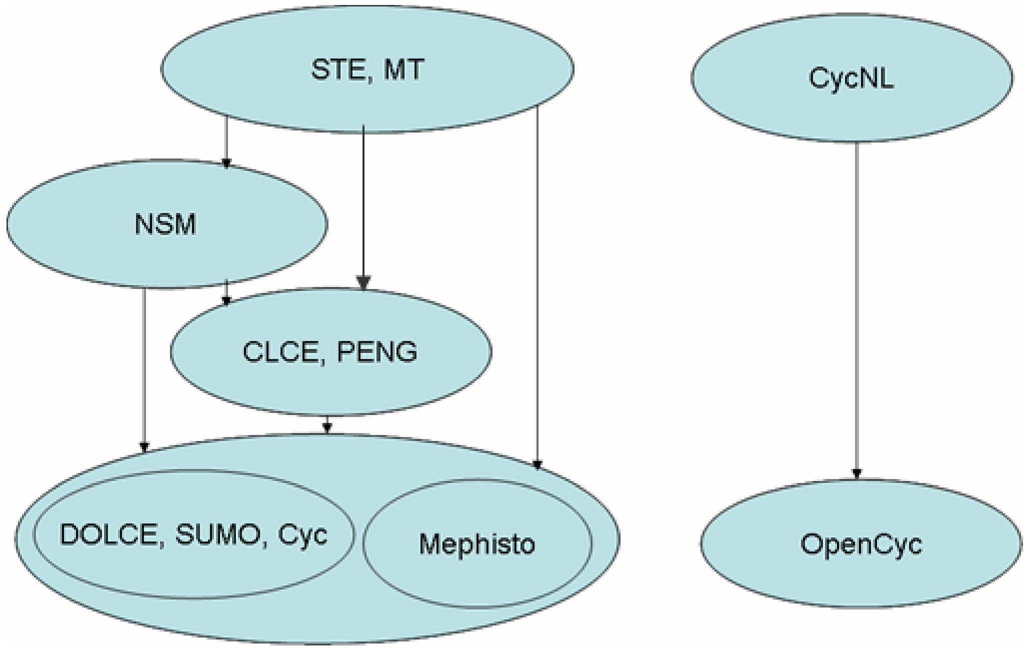

Figure 13: Options

The recommendation is to explore:

- STE or MT for the highest level natural language;
- Possibly NSM for intermediate natural language translation;
- Possibly PENG or CLCE for formal translation;
- One of DOLCE, SUMO and Cyc, plus Mephisto for the formal level;
- Consider the consequences for interfacing to JC3IEDM.

11. References

[1] D. A. Lambert, Situations for Situation Awareness, Proceedings of the 4th International Conference on International Fusion, Montreal, 2001.

[2] D. A. Lambert, Grand Challenges of Information Fusion, Proceedings of the 6th International Conference on Information Fusion, Cairns Australia, 2003, pp. 213 – 219.

[3] F. E. White, A Model for Data Fusion, Proceedings of the 1st National Symposium on Sensor Fusion, 1998.

[4] A. N. Steinberg, C. L. Bowman F. E. White, Revisions to the JDL Data Fusion Model, The Joint NATO/IRIS Conference, Quebec, 1998.

[5] A. N. Steinberg, C. L. Bowman, F. E. White, Revisions to the JDL Data Fusion Model, Proceedings of the SPIE, Vol 3719, 1999, pp. 330 – 441.

[6] A. N. Steinberg, C. L. Bowman, Rethinking the JDL Data Fusion Levels, Proceedings of the National Symposium on Sensor and Data Fusion, John Hopkins Applied Physics Laboratory, 2004.

[7] J. Llinas, C. Bowman, G. Rogova, A. Steinberg, E. Waltz, F. White, Revisiting the JDL Data Fusion Model II, Proceedings of the 7th International Conference on Information Fusion, Stockholm, Sweden, 2004.

[8] http://en.wikipedia.org/w/index.php?title=Controlled_natural_language&oldid=147504136.

[9] AeroSpace and Defence, Simplified Technical English, vol. 2007, AeroSpace and Defence, Brussels, 2007.

[10] J. F. Sowa, Common Logic Controlled English, http://www.jfsowa.com/clce/specs.htm, 2004.

[11] A. Wierzbicka, Semantics: Primes and Universals, Oxford University Press, Oxford, New York, 1996.

[12] A. S. Hornby, Oxford Advanced Learner's Dictionary of Current English, Oxford Univeristy Press, Oxford, 2005.

[13] W. Frawley, Linguistic Semantics, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1992.

[14] J. F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

[15] R. Jackendoff, Semantic Structures, The MIT Press, Cambridge, Massachusetts, 1990.

[16] Aristotle, Topics, in: W. D. Ross (Editor), The Works of Aristotle, vol. i, Clarendon Press, Oxford, 1937.

[17] B. Pascal, De l'esprit géométrique et de l'art de persuader, in: J. Chevalier (Editor), Œuvres complètes, Gallimard, Paris, 1667/1954, pp. 575-604.

[18] G. W. Leibniz, Opuscules et fragments inédits de Leibniz, Presses Universitaires de France. Repr. 1961 Hildesheim: Georg Olms, Paris, 1903.

[19] Z. Maalej, Review article: The Semantics of Polysemy, vol. 2007, LINGUIST List 17.106, 2006.

[20] L. Bruce, Review article: Emotions in crosslinguistic perspective, SIL Electronic Book Reviews 2004-009, SIL International, 2007.

[21] http://www.loa-cnr.it/DOLCE.html.

[22] S. Borgo, A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, Ontology Roadmap. Technical Report D15, WonderWeb Project Deliverable, 2002.

[23] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, L. Schneider, The WonderWeb Library of Foundational Ontologies. Technical Report D17, WonderWeb Project Deliverable, 2002.

[24] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, Ontology Library. Technical Report D18, WonderWeb Project Deliverable, 2003.

[25] http://en.wikipedia.org/wiki/Cyc#Criticisms_of_the_Cyc_Project

[26] http://www.cyc.com/cyc/technology/technology/whatiscyc_dir/whatsincyc

[27] http://www.cyc.com/doc/handbook/oe/oe-handbook-toc-opencyc.html

[28] http://www.cyc.com/cycdoc/upperont-diagram.html

[29] Nowak, C. (2003). "On ontologies for high-level information fusion". Proceedings of the 6th International Conference on Information Fusion. Cairns, Australia. pp. 657 - 664.

[30] Dennett, D. C. (1971). "Intentional Systems" Brainstorms : philosophical essays on mind and psychology. p. 3 - 22. Edited D. C. Dennett. The Harvester Press Limited, Brighton, Sussex.

[31] Lambert, D. A. (2008). "A Blueprint for Higher-Level Fusion Systems", *submitted to* The Journal of Information Fusion, Elsevier Press.

[32] Nowak, C. and D. A. Lambert (2005). "The Semantic Challenge for Situation Assessments", Proceedings of the 8th International Conference on Information Fusion, Philadelphia, U.S.A..

[33] I. Niles, A. Pease, Origins of the IEEE Standard Upper Ontology, Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology, Seattle, WA, USA, Vol. 3742, 2001.

[34] S. K. Semy, M. K. Pulvermacher, L. J. Obrst, Toward the Use of an Upper Ontology for U.S. Government and U.S. Military Domains. Technical Report, Mitre Corp., 2004.

[35] Ontology Portal Website http://www.ontologyportal.org/

[36] V. Mascardi, V. Cordi, P. Rosso, A Comparison of Upper Ontologies, Technical Report DISI-TR-06-21 (from the Dipartimento di Informatica e Scienze dell'Informazione, Università degli Studi di Genova), 2006.

[37] Ontology Portal Website - Publications http://www.ontologyportal.org/Pubs.html

[38] I. Niles, A. Pease, Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology, in The Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE '03), Las Vegas, NV, USA, 2003.

[39] I. Niles, Mapping WordNet to the SUMO Ontology, Teknowledge Technical Report, 2003.

[40] Ontology Portal Website – What's New http://www.ontologyportal.org/WhatsNew.html

[41] P. Baumgartner, F. M. Suchanek, Automated Reasoning Support for First-Order Ontologies, in The Proceedings of the Principles and Practice of Semantic Web Reasoning, 4th International Workshop, PPSWR 2006, Vol. 4187, pp. 18-32 of Lecture Notes in Computer Science, Budva, Montenegro, 10-11 June, 2006.

[42] A. Riazonov, A. Voronkov, Vampire 1.1 (system description), in The Proceedings of the International Joint Conference on Automated Reasoning, Vol. 2083 of Lecture Notes in Computer Science, Springer-Verlag, 2001.

[43] Common Criteria for Information Technology Security Evaluation Website http://www2.imm.dtu.dk/~robin/CC/

[44] D. J. Russomanno, C. R. Kothari, O. A. Thomas, Building a Sensor Ontology: A Practical Approach Leveraging ISO and OGC Models, Proceedings of The 2005 International Conference on Artificial Intelligence, Las Vegas, NV, USA, pp. 637-643, 2005.

[45] A. Pease, Tutorial Presentation: Introduction to Ontologies, Available from http://www.ontologyportal.org/pubs/ontolog-ontologyTutorial--AdamPease_20030529.ppt#270,15,Simple Methodology

[46] SUMO V.1.73 KIF Listing http://suo.ieee.org/SUO/SUMO/SUMO_173.kif

[47] P. Grenon, B. Smith, L. Goldberg. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In D. M. Pisanelli (ed.), Ontologies in Medicine, IOS Press, 2004.

[48] The KSMSA Website http://virtual.cvut.cz/ksmsaWeb/discussion/index.html?subject=ontology-org.ksmsa.ontologies%2Fsumo%23sumo

[49] http://www.ifomis.uni-saarland.de/bfo

[50] P. Grenon, BFO in a Nutshell: A Bi-categorial Axiomatization of BFO and Comparison with Dolce. IFOMIS Technical Report, 2003.

[51] P. Grenon. Spatio-temporality in Basic Formal Ontology. SNAP & SPAN, Upper-Level Ontology, Framework for Formalisation. IFOMIS Technical Report, 2003.

[52] A. Kumar, B. Smith. The Unified Medical Language System and the Gene Ontology. In A. Gunter, R. Kruse, B. Neumann (eds.), KI 2003: Advances in Artificial Intelligence (LNAI 2821), Springer, 2003.

[53] D. Oberle, Semantic Management of Middleware. In Proceedings of the 1st International Doctoral Symposium on Middleware. ACM Press, 2004.

[54] L. Schneider, Designing Foundational Ontologies – The Object-Centered High-Level Reference Ontology OCHRE as a Case Study, in: Song, I.-Y.; Liddle, S.W.; Ling, T.W.; Scheuermann, P. (eds.): Conceptual Modeling - ER 2003, 22nd International Conference on Conceptual Modeling, LNCS 2813, Heidelberg: Springer, 91-104, 2003.

[55] http://www.mip-site.org/

[56] http://www.mip-site.org/publicsite/03-Baseline_2.0/C2IEDM-C2_Information_Exchange_Data_Model/

[57] http://www.mip-site.org/publicsite/03-Baseline_2.0/C2IEDM-C2_Information_Exchange_Data_Model/C2IEDM-Annexes-UK-DMWG-Edition6.15e-2005-12-02.pdf, pK-1 (PDF document page number 525).

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)
An Overview of Conceptual Frameworks	Document (U) Title (U) Abstract (U)

4. AUTHOR(S)	5. CORPORATE AUTHOR
Dale Lambert, Adam Saulwick, Chris Nowak, Martin Oxenhan and Damien O'Dea	DSTO Defence Science and Technology Organisation PO Box 1500 Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-2163	AR-014-249	Technical Report	May 2008

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2007/1149188	07/149	CDS	60	57

13. DOWNGRADING/DELIMITING INSTRUCTIONS	14. RELEASE AUTHORITY
To be reviewed three years after date of publicationDSTO-TR-2163	Chief, Command, Control, Communications and Intelligence Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.shtml

Knowledge representation, Knowledge engineering, Data fusion, Information fusion

19. ABSTRACT
This document provides a short overview of the semantic frameworks that are applicable to information fusion applications. The frameworks considered include: STE, NSM, CLCE, DOLCE, OpenCYC, Mephisto, SUMO, BFO, and JC3IEDM.